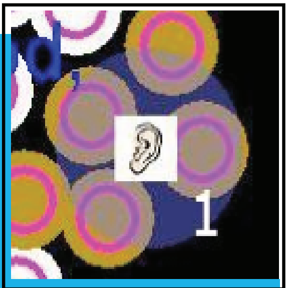
Supporting Seeking Tasks within Spoken Word Audio Collections

Lyon, Kirstin C.



Kirstin Lyon

Supporting Seeking Tasks within Spoken Word Audio Collections

Department of Medialogy
Aalborg University Esbjerg

AALBORG UNIVERSITY
ESBJERG

Supporting Seeking Tasks within Spoken Word Audio Collections

by

Kirstin Lyon

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Aalborg University
Aalborg University

Abstract

This thesis presents research that focuses on supporting seeking tasks within spoken word collections. Existing literature argues that different types of seeking tasks exist. Difficulties associated with finding relevant information are connected to the size of the collection, individual experience, how well a collection is known and the type of information in the collection. Most seeking studies typically focus on text and music information. The aim of this work is to investigate seeking behaviour within spoken word audio collections. Existing studies indicate that using visualisation techniques as a presentation method for collections may be an effective solution and support the different types of common seeking tasks. These techniques are applied to spoken word collections in this work.

Difficulties commonly associated with visualisation methods are that of visual overload and representing multi-dimensional information. Visual overload occurs when the amount of visual information present is too large for users to process. Multi-dimensional information collections are difficult to represent as the number of relationships within the collection can be highly complex making the visualisation difficult to create, and to navigate. Both difficulties can increase the time and error rate of finding relevant information.

In this research, Personal Audio Work Space (PAWS), and, later, the PAWS II interfaces, were designed with the aim of supporting seeking tasks within spoken word collections. The features of PAWS and PAWS II attempt to improve performance in seeking tasks by increasing the number of modalities and providing additional support for navigation. There is a particular emphasis on supporting and extending individuals' natural capabilities. PAWS and PAWS II were tested empirically, and the results are presented in this thesis. They indicate that the organisation of a collection is an important factor when performing seeking tasks. They also indicate that the simultaneous

presenting multiple sounds enhances seeking performance when the sought keywords are known.

Acknowledgements

Firstly I would like to thank Kristoffer Jensen for supporting me in my PhD. His comments, perspectives and thoughts have helped shape this thesis. I would not have reached the end without his support and advice. Next, I would like to thank Kasper Hornbæk for his time and advice concerning both of the experiments. Thanks also to the staff of the *Know Center* who not only provided two enjoyable summers, but also support, advice, and participants for the first experiment. Next, I would like to thank the various Java gurus who have helped me in implementing my prototype. To Erling Tind, for his programming assistance and excellent proof reading skills, and to Alice Niethammer for her careful proofreading of the final version. Thanks also to the staff of AUE, both past and present.

Finally, I would like to thank my family and friends. They not only participated in interviews and experiments, but also provided valuable friendship and support, especially to my Mum who bravely asked "what are you up to?" on a regular basis despite knowing the response was likely to be a rant!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Digital information collections, both personal and public, continue to increase both in size and diversity. The number of collections available to the general public is growing, with institutions such as the BBC [1, 2, 3] and Denmark's national broadcaster DR [4] making their radio and TV archives available to the general public, as well as the increase in digital libraries. A continuing self-publishing trend coupled with improved possibilities of sharing and creating information has meant that information spaces such as *MySpace* [5], *YouTube* [6], and *Flickr* [7] continue to grow in size and attract new users, both as creators and collectors of information. The way in which public and personal collections interact with each other indicates that the expansion and diversity of public and personal collections are linked. With the growing availability of collections as information resources, greater possibilities of storing personal information in digital form, and personal computers having more storage space [8], individuals have potentially large personal multi-media collections.

There are a number of reasons the increase in information collections has occurred. Firstly, people want to have information at their fingertips [9]. The use of online forums, discussion sites, blogs and websites are well-known methods individuals use to make decisions about from a range of choices, such as what products to buy, where to go on holiday, what company to work for and what education to pursue. Other forms of data searching include using websites to find details such as addresses, phone numbers, maps, and timetables. Secondly, people want to express their creativity and individuality by producing information, such as pictures, music and text and, in some cases, share their efforts with appreciative communities. Finally, technology for supporting content generation is cheap, easily available, and easy to use [10]. An implication of the improved technology is that now more people are able to contribute to information collections. Previously, information stored in a permanent form (e.g., CDs, books, and film) was produced by a relatively small group of individuals, such as musicians, writers, broadcasters, publishers, and researchers. This meant that there was a limit to how much permanent information could be produced, as relatively few people were creating it. Information was expensive to publish and all information was reviewed prior to release. Casual users acted as consumers of information, not creators. Now, casual users are becoming creators of information as well as consumers.

The growing amounts of information is associated with a number of difficulties. In 1945, Bush [11] stated that the amount of information was increasing in such a way that academics were, at that time, unable to read all of the material related to their field, and in some cases were unable to find all of it. He referred to this problem as information overload. Later, Woods et al. described three different ways in which people could experience data overload [12]: clutter, or just too much "stuff"; a bottleneck (i.e., not enough time to read available data); and knowing what information was relevant before a task was defined. Later, Denning [13] described a similar problem known as *infoglut*. *Infoglut* can be described as "cheap, digital information that is created faster than it can be processed by individuals". This means that, similar to information overload and data overload, people cannot keep up-to-date with all relevant information. *Infoglut* is part of a larger glut in communication technology where people are inundated with information, such as e-mail, voicemail, SMS messages, newspapers, and TV. Another difficulty associated with information overload is that of screening information. Not all information that is created is wanted. Examples of unwanted information include spam e-mail, pop-up advertising and malicious software. As a result of information overload, the time it takes to find relevant information can increase dramatically, while the number of relevant results found and the level of satisfaction with the results can decrease. The worse case scenario of not locating the right information at the right time is perhaps less important for casual users; however, businesses will suffer costs in both time and money, and for health care professionals and counter-terrorism failing to find the right information at the right time can be deadly.

1.1 Spoken Word as an Information Resource

Spoken word documents are information objects that contain spoken textual information, such as speeches and interviews, in their original audio form. In some cases spoken word documents include other forms of audio information, such as music and sound effects. The internal structure of a spoken word document can be highly organised, with clear introductions, sections and conclusions, it can consist of just free-flowing speech, or be a combination of the two. Spoken word documents come in a variety of topics and are used for entertainment, work, and education purposes. Different terms to describe spoken word information are used, including oral history [14, 15], spoken text [16], speech-as-data [17], speech [18], spoken audio documents [19], and spoken documents [20, 21, 22, 23]. For this thesis, all information that is presented in speech is referred to as a spoken word document, where a spoken word collection contains more than one spoken word document and may have either an implied or a stated organisation scheme.

There are a number of reasons the creation and accessing of spoken word documents as well as the number and types of user are increasing. For example, spoken word documents are easy, inexpensive, and quick to produce, making them a popular format for entertainment and for supporting education [24] and company training [25]. Creating spoken word documents is a popular way to capture tacit information [26], to support everyday memory [27, 8, 28], and to accurately record details from interviews, focus groups, and lectures. With Internet phone services, such as Skype [29], becoming a popular and inexpensive way to keep in touch with others, it is possible to record and store these conversations by using applications such as Kishkish Sam [30] and HotRecorder [31]. Following the self-publishing trend of blogging, audioblogging allows people to share their ideas through the audio medium [32]. In 2004, a group of audiobloggers began to use RSS technology [33] to allow casual users to automatically download the updated version of the audioblog that they had subscribed to. The combination of RSS technology with an audioblog is known as a podcast [34] (for videos, it is known as vodcasts). Recently, digital libraries consisting of collections of spoken word documents are being made accessible to the general public [14, 35, 4, 1]. With all of the spoken word collections available, and different uses for spoken word documents, people have more possibilities to create personal spoken word collections than before. This implies that understanding how people find spoken word documents and how to support these behaviours is an important factor in providing tools that support seeking within both public and personal spoken word collections.

Traditionally, the difficulties of using spoken word documents have been in accessing the information once created [36], scanning it, and extracting relevant information from it [37]. The difficulties are due partly to the transient and temporal nature of audio [36]. This implies that, a spoken word document takes longer to process and to search than a written text because a user cannot determine whether it contains the "right" information before listening to it. Additionally, a user must follow the pace dictated by the speaker of the document. In order to improve accessibility and ease of use of spoken word documents, a number of strategies have been suggested. Firstly, to transcribe spoken word documents to text and then to use text-based tools to extract information. Secondly, to preserve the spoken word document in its audio form and provide better tools for accessing the audio content. Thirdly, a hybrid of the two approaches.

In many cases spoken word documents are transcribed to text, with the audio record either no longer available or stored in a different place than the text [15]. A reason for this strategy could be a lack of tools allowing users to work effectively with spoken word documents in their original audio format. When spoken word documents are transcribed to text, text-based tools may be used to explore the information. This has the advantage that the time spent to locate wanted information may be reduced and that many tools exist to support seeking tasks in text-based information. This approach has a number of difficulties. Firstly, transcription is a long, tedious, and error-prone task. A one-hour recording takes approximately six hours to transcribe [22]. One approach to reduce the time to transcribe speech is to use automatic speech recognition (ASR) techniques, which automatically transcribe spoken word to text [38]. However, this is difficult to perfect for a number of reasons, such as: tone of speaker, dialect, accent, rate of speech, and number of speakers [39]. Also, representing a spoken word document visually means that certain data will be lost during the transcription process. Examples difficulties are how to represent a pause, and what to include in the text when more than one person is speaking at one time.

Recently, arguments for retaining spoken word in its original form have been made by different groups of users, such as oral historians [15] who consider spoken word documents to be historical artefacts, casual users who prefer listening to the documents, than reading them, and those who are either not able to read (i.e., the visually impaired), or do not want to read (i.e., users who are on the move). Tools have been created that aim to support better access to spoken word collections in their original audio form. One type of tool combines ASR techniques with text queries. For example, users enter a text query, this query is matched against a spoken word collection using ASR, and relevant titles of results are then displayed in lists. Users select a title, and the original spoken word document is produced, not a text transcript of the spoken word document. Example tools are TVEyes [40] and Podzinger [41]. Other tools that improve access to the audio content within a spoken word document are SpeechSkimmer [36] and DiMaß [42]. These tools introduce different audio processing techniques that speed-up the audio playback without reducing the intelligibility of the information.

The third approach is to use a hybrid of the spoken word document and its corresponding text transcription. Research shows that many users of spoken word collections take notes to aid them [37], which even though incomplete, proved helpful when searching for relevant information. With this in mind, collections such as the Shoah Foundation [14] and the National Gallery of Spoken Word (NGSW) [35] combine audio content with partial transcripts that can be searched. In this way, they combine the benefits of being able to scan text with the richness of information offered by audio recordings. However, it is not certain that all user groups would create a partial transcript of their whole collection unless the perceived benefit was high.

1.2 Thesis Aims

The motivation of this work came from a combination of personal experiences of transcribing interviews and focus groups for analysis purposes, a discussion with the *Dansk Folkemindesamling* [43], and the observation that a potentially valuable information resource was being overlooked, possibly due to the difficulties of accessing the information. A number of user types frequently use spoken word collections, and recently, with the increase in spoken word collections in the form of podcasts, a new generation of users is becoming interested in this type of information collection. Improving the accessibility of these collections should increase their use and hopefully enable both the original and new user groups to gain better access to spoken word collections.

The overall aim of this thesis is to better understand how seeking tasks in personal spoken word collections may be best supported while retaining spoken word documents in their original audio format. This work concentrates on seeking tasks within spoken word collections rather than individual spoken word documents. This means that the focus is on finding the right spoken word document within a collection, not the right information within a spoken word document. There is also a particular emphasis in this work on supporting user behaviour rather than replicating it. This investigation is intended to provide results that reduce the time it takes to locate a relevant spoken word document(s), that increase the chances of finding the right spoken word document(s), and improve user satisfaction while seeking within spoken word collections. These results can be used to improve the design of seeking tools and make spoken word collections more attractive for its many potential users. There is a particular emphasis on personal and small collections, but it is expected that the results may be applicable for larger, public collections as well.

The research can be divided into two general areas. Firstly, what seeking tasks and strategies should to be supported, and secondly, how to support these common seeking tasks and strategies. These areas and relevant research questions, as well as the research methodology for each question, are presented below.

1.2.1 What Seeking Tasks and Strategies Need to be Supported?

The focus of this research area is on understanding how people interact with spoken word collections. The following questions will be discussed for this research area:

RQ1 Identify possible seeking tasks relevant to spoken word collections.

RQ2 Identify possible seeking strategies frequently undertaken when searching spoken word information.

In order to answer these questions, literature from the fields of information science and information interaction is reviewed. This provides insights into the types of seeking task that are conducted, into what strategies and techniques are employed, and what sub-tasks are frequently performed in order to aid users in finding relevant information. This analysis is augmented by a study where four groups of users are interviewed about their seeking habits within their personal music collections. The results are analysed using a grounded theory approach and are combined with the initial review to provide a list of tasks, strategies and sub-tasks that will be considered for spoken word collections. Furthermore, results obtained from two empirical experiments, where spoken word collections are used, attempt to give further insight into whether the seeking behaviours and strategies described for music and text collections are also applicable to spoken word collections.

1.2.2 How To Support These Common Seeking Tasks and Strategies

The second research area focuses on how technology may be used to support the tasks and strategies outlined in RQ1 and RQ2. For this work, the research area is divided into the following questions:

RQ3 How should a spoken word collection be organised?

RQ4 What features support the different seeking tasks and strategies?

RQ5 How can the audio medium be used to best effect within a tool?

Using the initial results obtained in RQ1 and RQ2, as a starting point, a review of tools that focus on these results provides recommendations on how to present a spoken word collection and what features should be included in a tool, as well as the role of audio, in order to partly answer RQ3, RQ4 and RQ5. Personal Audio Work Space (PAWS) is designed using this analysis and an empirical study using 18 participants is performed using this tool. A statistical analysis of the results aims to show what features improve and hinder performance within two types of seeking tasks. Using these results, Personal Audio Work Space II (PAWS II) is designed and tested empirically using a different set of 18 participants. For this experiment, the participants complete four types of seeking task are completed by participants. The data obtained are tested statistically. The recommendations from both tests aim to answer RQ3, RQ4 and RQ5 more completely.

1.3 Thesis structure

Chapter 2 provides an overview of seeking tasks and strategies in both personal and public collections. After an initial review of prior work in the areas of information science and information interaction, a study involving four types of user is described. This chapter provides a list of common seeking tasks and expected strategies that may be applicable to spoken word collections and aims to partly answer RQ1 and RQ2, as well as providing suggestions for RQ3.

Chapter 3 demonstrates how the tasks described in RQ1 and RQ2 are being supported, and what aims still need to be supported. The chapter is organised based on the three classifications of users identified in Chapter 2. It concludes with an overview of tools and techniques that have been developed for audio information and use audio as the main form of presentation. This chapter suggests a possible presentation method and other features in order to partly answer RQ3, RQ4 and RQ5.

Chapter 4 presents the organisation and re-find experiment. It begins by describing PAWS (Personal Audio Work Space) that was designed using evidence and ideas gathered in Chapters 2 and 3. The aim of PAWS is to aid manual organisation and re-finding tasks within personal spoken word information collections by supporting frequently reported behaviours. The obtained results are tested statistically and aim to show what features have the most potential.

Chapter 5 presents the seeking tasks experiment. It begins by describing PAWS II, based on PAWS but with improvements based on the findings of the organisation and re-find experiment. For this experiment, participants are required to perform four types of seeking task using an automatically organised collection. The obtained results are a combination of qualitative and quantitative data. The experiment aims to provide a better understanding of how users seek within an unknown collection.

Chapter 6 summarises the work of this thesis and demonstrates to what extent the research questions in Section 1.2 have been answered. It explains the limitations of the work and how the limitations were balanced. The chapter concludes by providing suggestions for future work.

Chapter 2

Seeking in Personal and Public collections

When working with information collections, an aim for individuals is to be able to find relevant information efficiently and effectively. Individuals employ different strategies to aid them in this goal. There is an indication that the adopted strategies depend on the information medium (e.g., text, music, or video), what the end goals are, how well users can articulate what they want to find, collection organisation, and user preference. Most seeking research focuses on seeking text information in both personal (e.g., [44, 45, 46]) and public collections (e.g., [47, 48, 49]). Although this research is applicable to spoken word collections, it may not take into account the difficulties that seeking within audio collections presents. At present, seeking task research within audio collections appears to focus, for the most part, on music collections. An advantage of analysing seeking tasks and strategies within music collections is that it can give some insights into how users seek within audio information collections. For seeking within text collections users have developed strategies that have developed over time, and continue to evolve. It may be assumed that effective strategies for seeking within audio collections may have developed, and be different to those of text-based seeking strategies. Differences between spoken word, music, and sound sample information exist. Spoken word documents are easier to transform into a visual form than both music or sound samples (see Sect. 1.1 for examples of accessing spoken word content). The contents of spoken word documents are less subjective than those of music or sound samples, which implies that seeking within spoken word collections offers different challenges to that of seeking within music or sound sample collections. At present, research in the area of audio collections focuses on content extraction techniques [42]. Similarly, studies on music information retrieval and music digital libraries, focus on feature extraction. Only little empirical evidence for audio collections exists that focuses on user seeking styles and strategies [42]. The aim of this chapter is to propose what kind of seeking tasks and strategies may be used in connection with spoken word collections. To do this, the chapter begins by giving an overview of seeking tasks and strategies from web studies and information science literature. It then applies this knowledge to personal music collections by combining seeking behaviour studies and a study presented as part of this thesis using casual and professional users with both physical and electronic personal collections. It concludes with a list of seeking tasks and strategies that may be used within spoken word collections, and an overview of organisation styles and behaviours that may be employed.

2.1 Overview of Seeking Tasks and Strategies

This section begins by providing an overview of the seeking process. Next it provides an overview of the different information types and collections, including public and personal ones and how collections may be organised. Finally, it describes navigation strategies, seeking tasks, and strategies outlined within text-based studies.

2.1.1 Models of Seeking

Before seeking, users have an information need. That is, “the perceived need for information that leads to someone using an information retrieval system in the first place” [50]. Users come to perceive such a need for various reasons. Web studies that describe different reasons for seeking information include studies by Broder, Jarvelin et al, and kellar et al [49, 51, 52]. Once an information need has been perceived, users employ different strategies to satisfy that need. Wilson [53] proposed a user studies model that attempts to describe the concepts and relationships within information science that are applicable when seeking information. He states that seeking information is complex and has many factors. There is also a possibility of success and of failure (see Fig. 2.1 for an overview). Once an information need is formed, users employ some form of seeking behaviour, either by using an information system or information sources, or by discussing their information needs with other people. Each of these methods may be successful or fail. The success of finding wanted information may lead to an information use which can then either be transferred to other users or used to refine and information seeking need.

Traditional information retrieval (IR) methods focus on query matching between users' queries and a database of information (see Fig. 2.2) [54]. This means that the information need described by Wilson is formulated into a question (query) that is then compared against a collection. If the posed question matches any information within the collection, the result(s) will be displayed. Traditional IR models are limited in that users need to form queries that a machine can interpret, instead of using natural language. This model does not take into account the evolving nature of a search [55]. The IR model was extended by Broder to show how users seek information (see Fig. 2.3) [49]. First, users are faced with a particular task. From this, an information need may occur. In order to create a query, users must create a verbal formulation. This could be either speech or text. The query is entered into a collection's search engine, which then compares it with the complete works contained within that collection and returns a result(s). Users either

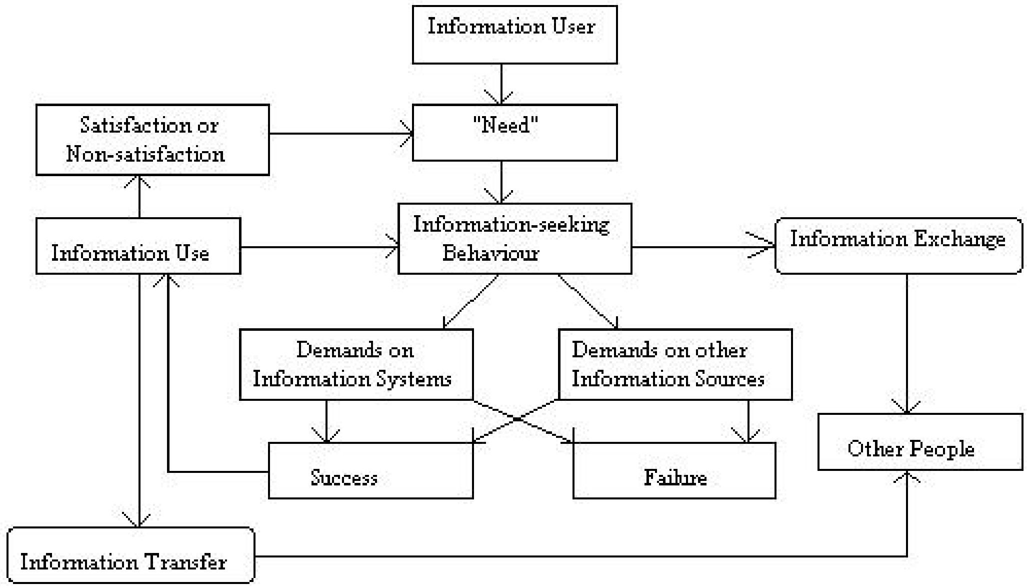


Figure 2.1: Information behaviour model proposed by Wilson that aims to represent the complexity of information seeking.

refine their query and continue to search, or stop. This implies that seeking is often an iterative and evolving process.

An example of an evolving or iterative seeking model is Berrypicking, which is perhaps a more accurate description of how people seek in electronic environments [55]. One difference between this and the traditional IR model is that the Berrypicking model assumes that users' information needs evolve depending on what information they find. The nature of the query as well as the nature of the overall search process evolves during the query time (see Fig. 2.4).

2.1.2 Collections

Another factor that affects how easily users locate wanted information is how large, diverse, organised and how well-known to users the collection is. Public collections are collections that are potentially accessible to all users, such as the World Wide Web and

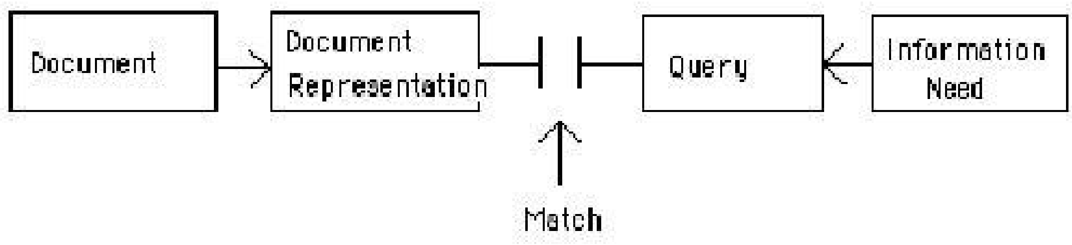


Figure 2.2: Traditional IR view of a seeking task. Queries are matched against a collection with results being returned.

digital libraries. Personal collections differ to public collections in many respects, such as they are usually smaller, assembled by a single user, contain information that reflect the wishes of a single user, and their developers are experts of the collection. In this case an expert is someone who's level of familiarity with a collection is greater than others. Personal collections are also less complete than public collections, due to factors such as the collector's preferences, available storage space and availability of electronic materials. Personal information collections differ from general information management in that professionals (e.g., librarians) do not manage the information [56]. In general, when an information need is first perceived, users look to their personal collections first before seeking within a public collection [57].

Organisation of Collections

A factor that influences how people navigate and seek within collections, is how a collection is organised. Collections, whether personal or public, can be organised in a number of ways. Typically, public collections are categorised in some way, whereas personal collection are arranged as decided by their owners. An example of a highly structured and complete public collections are digital libraries. Whereas, the Web [58] would be considered to be a semi-structured public collection. A difference between the two is that a group of experts has organised the first collection anticipating how it will grow. The

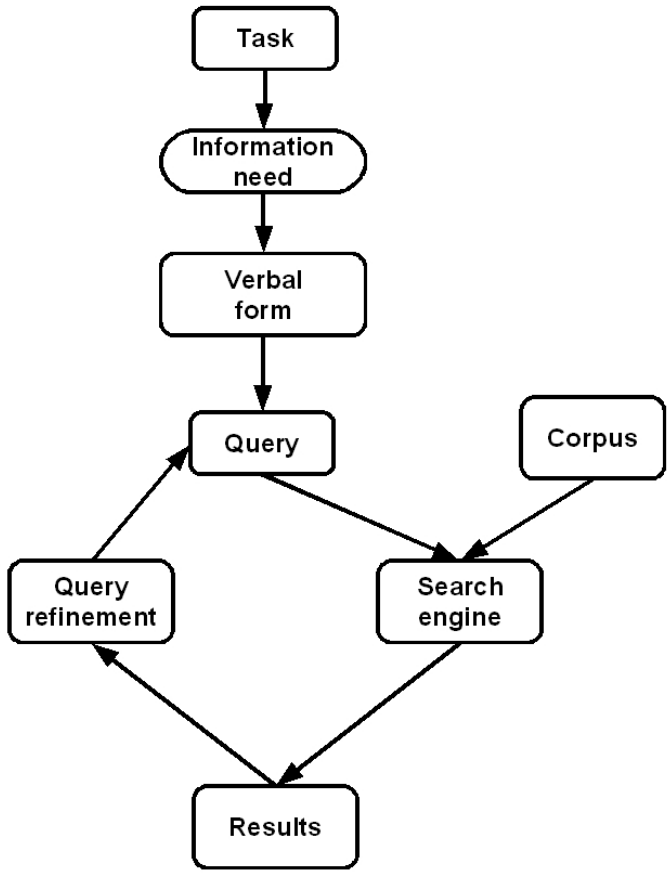


Figure 2.3: Traditional view of IR seeking task extended by Broder.

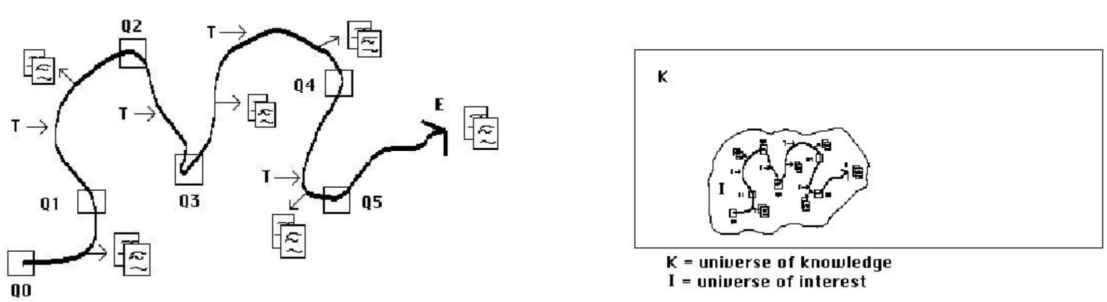


Figure 2.4: Berrypicking model of seeking information on the Web.

Web is organised by its users, who can be inconsistent and do not need to understand a complex structure or how it will grow in order to participate in it.

Organising information is a non-trivial problem. Different factors include: the number of people structuring a collection, its size, diversity, and homogeneous, and whether the goal of a collection is known and well-defined prior to structuring. The structure of collections may change over time, when the initial system no longer makes sense to its creator. Other difficulties include where to place information that either fits into more than one place, or has no place at all.

Information architecture describes two categories of organisation schemes: exact and ambiguous organisation schemes [59]. Exact organisation schemes divide information into well-defined and mutually exclusive sections. Examples of exact organisation schemes are alphabetical, chronological and geographical schemes. Ambiguous organisation schemes divide information into categories that defy exact definition. Issues of ambiguity in language and organisation, as well as user subjectivity, mean that these schemes can be difficult to design and maintain, but tend to be more useful than exact organisation schemes. Typical schemes include topic, task, audience, metaphor and hybrids schemes.

Other organisation structures that are commonly used are hierarchy and hypertext structures. A hierarchy structure uses files and folders to describe the relationships within a collection, and is sometimes known as a classification or file/folder. A classification scheme can be defined as "a spatial, temporal or spatio-temporal segmentation of the world" [60]. Bowker and Star propose that there are three theoretical properties to a classification scheme. First, a consistent, unique classificatory principle, such as temporal or alphabetical ordering. Secondly, mutually exclusive categories, and thirdly, completeness. They also suggest that adhering to these theoretical principles is not always possible in real world classifications. Bälter [61] described the retrieval benefits of organising by folders to be a reduction of the search space. Instead of scanning a long list of unstructured items, users can home in on a particular category. Another benefit of the file/folder system is that it allows people to deal with the complexity of systems [62]. There are a number of limitations to such systems. A fundamental flaw is single inheritance [63, 64], this means, how to position information that can belong in more than one place. Multiple links are possible, but often confuse users [64]. File/folder systems are also criticised for their static nature and poor scalability of large information spaces [65].

Hypertext is a highly non-linear method of structuring information. A hypertext system involves two types of components: information objects, and the relationships between those information objects. Individuals decide on the relationships between information, making hypertext highly flexible. However, as many of the relationships describe personal associations, users may experience navigation difficulties. A structure being discussed within the hypertext and visualisation communities is spatial structures or organisations. Spatial organisation occurs when the relationships between information within a collection are described using visual properties, such as proximity, colour, size and spatial location. Spatial organisations take advantage of users' perceptual abilities to scan, recognise and recall images rapidly [66]. Users are able to automatically detect patterns and changes in colour, size, shape, movement and texture. Spatial organisations move the work from the cognitive to the users' visual perceptual system [66]. Other advantages of spatial organisations include its ability to facilitate constructive ambiguity, to support emerging problem solving strategies and to reduce the overhead in communicating to others [67]. Spatial organisations can be used for different occasions [66]. This ranges from a large collection that contains information outside the users environment to visual objects representing a small collection of users' materials. Disadvantages of spatial structures includes the difficulty of interpreting changes of organisation, organisations interfering with visual source information, inability to express cross-space inter-relations, and a lack of means for expressing publishable information spaces [68].

Collections may be organised either by users (user-centred), automatically, using algorithms (computer-centred), or a combination of the two approaches. There are a

	User-centred	**Computer-centred**
Organisation	Association, spatial	Classification
Task type	Complex, unstructured, long-lived	Repetitive, automatic, structured short-lived, on-the-fly
Emphasis	Remembering connections	Defining connections
Rules	Flexible	Well-defined

Table 2.1: Differences between organisations that are created by users and computers.

number of differences between manual and automatic organisations. Firstly, people and computers do not organise their information in the same way. People tend to organise by building associations between information or to organise spatially, whereas computer algorithms construct classifications. Another difference is that automatic organisation usually requires well-defined rules prior to organising information, whereas people are able to change the organisation rules to fit a given scenario. Computers are able to complete tasks that are automatic in nature more effectively than people, but are unable to organise information without prior rules. The main differences between user- and computer-centred organisation tools are described in Table 2.1.

A user-centred or computer-centred approach may be used to support seeking tasks. The approach of this research is to focus on tools that support users' natural capabilities. This involves techniques to strengthen memory cues, navigation that is similar to the physical world, and easy access to the audio content. An advantage of user-centred support is that they leverage skills that users already have. User-centred tools are being used successfully within the knowledge management community for tools that support knowledge work and decision making. These tools automate tasks that are well-defined (e.g., putting a list into alphabetical order), while ill-defined tasks are completed by users. It is expected that not all tasks and users will be suited to user-centred approaches.

Organisation Strategies within Personal Collections

An aspect that is discussed within personal information management research is how users organise and re-find information. There are three main issues when organising a collection.

- Should the collection be organised?
- How should the collection be organised?
- Will the original organisation be maintained?

Different studies have defined types of users by whether or when they choose to organise or maintain their collections. Whittaker et al. [69] found, in a study of e-mail organisation, that there were different types of users when it came to how a collection was organised and maintained. They described the users as filers. The types of filers they observed were: frequent filers, spring cleaners, and no filers. Frequent filers are those that maintain their organisations on a regular basis. Spring cleaners would re-organise their collection after a period of time when it was deemed the organisation was in need of a clean-up. No filers do not create or maintain any organisation scheme. Studies by Berlin et al. [70] showed that when dealing with electronic documents two kinds of personal strategy was apparent. Some people organised up front, others would rather spend time on retrieval than organisation.

A study by Malone [71] looked at how users organised their paper information. He proposed that the desktop metaphor of files and folders forced users to work in a way that was not natural to them; that is, users had to organise hierarchically. Later, Boardman et al [72] demonstrated that the behaviour Malone had described for the paper office was also true for electronic documents. When documents are in a folder the situation of "out of sight, out of mind"could occur [73].

Essentially, the perceived use of information determines where it will be placed [46]. The categories used to organise depend on the intended use. If the perceived use of information is unknown, organisation becomes more challenging [74]. In this case the cost of organising may be considered too high, with users choosing not to organise or maintain. Observed behaviour for organising information in both physical and electronic collections was shown to be consistent [75, 76].

Information Type

Information type also played a role in how/if a collection will be organised. There are a number of ways of describing information. For the purpose of this work, three descriptions are considered relevant: medium, frequency of access or use, and relationships between different information objects in collections.

Medium refers to what form the information is presented in. Examples of media are vision, sound, proprioception, touch, olfaction and gustation, and sensory integration and motor coordination [77]. This work focuses on information presented in the sound and vision media. Examples of information presented in vision information include text, graphics, and pictures. Information within the sound domain include music, speech, and sound effects. Naturally, both media may be combined for video information. Further information on media taxonomies can be found in a study by Heller et al [78].

Describing information by the frequency of its use was first outlined using paper documents. Three information types were first noted by Cole [79]: ephemeral, working, and archived. Ephemeral information typically has a short shelf live, consisting of e-mails, to do lists, etc. Working information is information that is used on a regular basis and is relevant to a user's working tasks. Archived information typically has a long shelf life (months to years) but offers little regular benefit to users' tasks. Nardi et al. [80] demonstrated, by observing office workers, that the information types described by Cole were also apparent in electronic collections. That information can be categorised into particular types depending on the frequency and type of its use has also been observed for other electronic information forms, such as documents [80, 75, 81] and e-mail [82].

	Cole (1982)	**Hertzum (1999)**	**Gwizdka (2000)**	**Cunningham et al. (2004)**
Frequently accessed	Ephemeral	Action information	Prospective & ephemeral	Active items
Occassionaly accessed	Working	Personal work files	Working	Occasional use
Seldom/never accessed	Archived	Archive storage	Retrospective	Archived

Table 2.2: Information types identified in personal information management (PIM) research organised by the perceived frequency of use of information.

Cunningham et al's [83] findings show that information types are also apparent within physical music collections. The type information has depends on its perceived or intended use [46] and is decided upon by its user(s). Table 2.2 shows four ways of defining information type by frequency of use.

Different ways to describe information regarding the relationships between information within a collection includes: 1D, 2D, 3D, multi-D, temporal, tree, and networks [84]. Table 2.3 provides an overview of each information type.

2.1.3 Navigation Strategies

In order to successfully seek information, users use techniques to navigate around information collections. In this case, navigation within electronic worlds has similar properties to navigation within physical worlds. Not surprisingly, the style of navigation chosen is dependent on the seeking task and user preference.

Web seeking research has found that different strategies have been developed by users. Two types of strategy are orienteering and teleporting [85]. Orienteering occurs when users use small jumps to find the desired information, and is an iterative approach. Teleporting refers to the users' strategy of moving to the wanted information immediately. Herbert et al. described other styles of browsing as flimsy and laborious [86]. This referred

Information Type	Description
1D (1-dimensional)	Information is linear (e.g., alphabetical list, sequential organisation)
2D (2-dimensional)	Sometimes known as spatial (e.g. maps, floorplans)
3D (3-dimensional)	Similar to real world objects, have volume, relationships to other objects
Multi-dimensional	Meta-data attributes such as type, size, author, modification date
Temporal	Information organised based on time
Tree	Information in a tree structure
Network	Information organised as a graph

Table 2.3: Description of information types using relationship as the main descriptor.

to how deep the navigation was. They used details such as path length, relative amount of re-visits, page return rate, back button usage, and relative amount of home page visits to gauge what style was used for different seeking tasks. To determine navigation complexity, they looked at the number of links followed per day, number of cycles, path density, compactness, and average connected distance.

Catledge et al. [48] found that users rarely traverse more than two layers before returning to an entry point, and that only 3% of pages were archived. This means that other navigation techniques must be employed in order to successfully navigate around a collection. Therefore, relevant information must be within two or three jumps from a users starting point. This means more cluttered information spaces increase the search time but also increase users chances of finding the right thing in the end. They also described the need for personalised views for individuals. Even though patterns exist in seeking, users' still find techniques that suit themselves [87].

Most studies focus on navigation within web-based electronic text worlds. Laplante et al. [88] found that being able to access the audio content easily and swiftly for music browsing was a high priority for many people. At present, research that attempts to understand listening behaviours is not available. Relevant questions on listening behaviours include how long is the audio content is listened to before a decision is made on whether it is relevant.

2.1.4 Seeking Tasks

Even though seeking in collections is an everyday activity, it is a non-trivial task. The difficulty of the task depends on factors such as user experience, domain knowledge, how well a task is defined, and how well a collection is known [47]. In studies by Bystroem et al. [47] and Kellar et al. [52] they showed that if users perceived the seeking task to be too difficult, they were likely to discuss their information needs with other users first, instead of using automatic tools. This implies that present tools are not able to support difficult-to-articulate information seeking needs.

Different seeking tasks exist. Tasks are chosen depending on factors such as how well an information need can be described and if the seeking task is a goal or non-goal (e.g., an example of a non-goal seeking task is when users "surf on the web". No real information need is formed, but information is found anyway). Seeking can occur in public collections where the information collection and/or information being sought is new or unknown to users, or in personal and public collections where the information is known or has already been found, and users wish to find the information again (re-find).

There are a number of differences between finding and re-finding information. Finding is concerned with looking for information for the first time and differs from re-finding in a number of ways [45]. An overview of these differences is provided in Table 2.4.

A research area that discusses re-finding within personal information collections as a sub-activity is personal information management (PIM). Topics that PIM is concerned with are the collection, storage, organization and retrieval of digital objects (e.g., files, addresses, and bookmarks) by an individual in their personal computing environment [89].

Finding	**Re-finding**
Uncertainty: Is the information available at all? Do I know the right keyword to use in a search engine?	Certainty: I have seen the information already, but where? What was the keyword that I used in the search engine?
Recognition: Is this the information I am searching for?	Recognition & Recall: Where did I see that? Is this where I saw it? Context is very important
Strategies: Intuition, search tools, foraging	Strategies: Directed (focused) navigation

Table 2.4: A comparison of finding and re-finding information by Capra et al.

Most research in the area of PIM discusses ways in which users re-find information from their visual collections, such as documents [75], webpages [73, 45, 90], e-mails [82], and pictures [91]. Examples of personal spoken word collections are podcasts [34], recorded telephone conversations [29], voicemail [92] and spoken word collections gathered from interviews (either from online collections [1], or created by users).

Both finding and re-finding have two broad categories of seeking tasks that can be described as searching and browsing. Searching is usually defined as a seeking task where the information need is known and can be articulated, whereas browsing seeking tasks occur when people are not able to articulate what it is that they are looking for, but can recognise relevant information when it is found. Various researchers have attempted to define seeking tasks according to different criteria including browsing and seeking, and whether the task is a goal or non-goal. These tasks can be further refined in a number of ways.

Aguilar et al [93] outlined a model for scanning information. In this sense, the use of the term scanning within information science is roughly comparable to seeking within web studies. They identified four types of seeking: undirected browsing, conditioned browsing, informal searching, and formal searching. Undirected browsing occurs when users have a general idea of what they are looking for, but are unsure what specific area they are specifically looking for. Conditioned browsing occurs when users know exactly what areas of interest they are looking for within a collection. Informal searching is where basic queries can be formed and a "good enough" result is sufficient. Finally, formal search occurs when users know precisely what they are looking for.

Marchionini [94] observed that there were three general types of browsing: directed browsing, semi-directed browsing and undirected browsing. Directed browsing occurs when the browsing is focused, systematic and directed by a specific object or target (e.g., scanning a list for a known-item). Semi-directed browsing occurs generally purposeful or can be predicted. A less systematic method used is, for example, entering general keywords into a search engine and examining the retrieved records. Undirected browsing occurs when there is no real goal and very little focus (e.g., "channel hopping").

Similarly, Wilson [95] identified four categories of information seeking and acquisition: passive attention, passive searching, active searching and ongoing search. Passive attention occurs when there may be no intention of seeking, but information is acquired anyway. Passive searching occurs when one type of search causes relevant information to be found. Active searching occurs when individuals actively seek out information and, finally, ongoing searching occurs in order to update or expand on found information.

There are a number of ways to describe seeking tasks and behaviours used within both personal and public collections, but for this work they can be described using the following questions.

- Am I actively seeking information? (passive vs. active)
- Have I found this before? (find vs. re-find)
- Can I articulate what I'm looking for in a way a search engine can understand it? (browsing vs. searching)
- Is there an exact answer? (best fit vs. acceptable)

- How much information do I need/want to find? (1 — many)
- What constraints do I have? (e.g., time, money, ...)

2.1.5 Seeking Strategies

Different strategies are used in connection to the various seeking tasks. Strategies for seeking information may be different depending on how the above questions are answered.

Studies by Choo et al. [96] outlined tactics or behaviours that were commonly used in order to complete the different seeking tasks outlined by Aguilare et al. They described four tactics that were used in seeking information: visioning, discriminating, satisficing and optimizing. Visioning is the behaviour of scanning a diversity of sources, taking advantage of easily accessible information. Discriminating is when users browse within pre-selected sources on pre-defined topics of interest. Satisficing is used when the search is focused on an area or topic, and a "good-enough" search is satisfactory. Finally, optimizing is a systematic gathering of information about an entity, following a particular method or procedure.

Other ways in which to describe seeking strategies are: location-based searching, linear searching and query searching.

Location-based searching involves looking at where the information is expected to be first [75]. This makes the organisation of a collection an important factor. A good initial first placement of information is important to the success of re-finding it again later. The location of an information object gives additional implicit information and acts as a reminder as to where a particular object is. Location and browsing relies on structure (prior organisation). Research by Barreau et al. [75] demonstrated that the most popular way that users searched for information was by location. The reason for this preference was its reminding function. Another possible reason for the popularity of location-based re-finding is due to the fact that this skill has been developing for many years using physical objects. The relatively recently available technique of searching by keyword is new. Barreau et al. found that it was how information was used that determined how it would be organised, and that retrieving and reminding were connected. Studies by Bruce et al. [97] show that when asked to recall features of their own information, users were successful in remembering the location of a file.

Linear searching involves starting at the beginning of a list and working through it until either the right object has been found or the end of the list is reached. It is a standard algorithm within Computer Science. This approach is used when users do not have a clear starting point, or have found an appropriate place that still requires further searching. It is typically a lengthy method of searching that requires patience.

Query re-finding involves entering keywords (or queries) into a tool. The tool matches the keywords with an information collection, and displays results that match the entered query. This approach is not available for physical objects. Bruce et al. [97] found that recalling the title or file name of a file was more difficult for participants than other attributes. Query re-find using audio properties (e.g., re-find a male voice) is difficult to achieve. This type of re-find technique was used more in the case where the title was known. For those users who used tools to organise their collection, it was more successful than those users who used no prior organisation.

2.2 Methodology

The aim of this study is to determine what seeking tasks are used within audio collections, and if the tasks and strategies change due to the user being either a professional or casual user, or if the collection is physical or electronic. To gather data for this study, four cases were considered. These cases combined casual and professional users with physical and electronic collections and were thought to cover typical users. Each interview was semi-structured with initial questions concerning the participants experience with organisation and re-finding within their personal collections. Prior work has discussed personal collections, but few have considered professional personal collections. It is expected that this knowledge can be extended into spoken word collections.

2.2.1 Overview of Interviews

Each participant was interviewed concerning their organisation and re-finding strategies. Notes were taken for each participant, and the duration of each interview was between 30 and 60 minutes. The results were analysed using Grounded Theory methods.

All participants fell within the ages of 20 to 40 years. The difference between casual and professional participants was that professional participants used their music collections for work purposes (i.e., the professional participants interviewed were three DJ's and a composer). A few of the casual participants worked in the music industry, but their music collections were for private use, not work; therefore, they are considered to be casual participants for this study.

Case 1: Casual Users - Electronic Music Collections

Twenty-four persons were interviewed. Of the 24 one had no collection of any kind, preferring to use the radio for listening to music. The size of the participants' collections varied from a few hundred megabytes to 60 gigabytes. The diversity of the music collections also ranged from a fairly homogeneous group, to a vast mixture of styles (highest was around 100 genres). Some participants' collections had been developed over a long period of time (years), whereas others had been developed in a short period, for example, they had transferred their physical CD collection to a digital format.

Case 2: Casual Users - Physical Music Collections

Seven persons were interviewed about their physical music collections (primarily CDs). The collections ranged in size from around 50 to a few hundred CDs. Diversity also ranged from a few genres to around 20. Most CDs are of a standard size, although some participants had box sets that could not be placed close to their other CDs. One of the participants had moved all the CDs from their cases to CD wallets in order to reduce the amount of physical space they took up. Another participant had a small collection of CDs at work.

Case 3: Professional Users - Electronic Music Collection

The composer had been writing music for approximately 15 years using mainly electronic methods (i.e., using Nuendo [98] to compose, not paper transcripts). His audio collection consisted of both virtual instruments and samples and amounted to around 250 gigabytes. All of the samples and virtual instruments had been bought from musicians specialising in sample production.

Case 4: Professional Users - Physical Music Collections

The three DJs were interviewed about the physical music collections they used for work purposes. All had been working professionally for at least two years, but had been interested in music before that. All of the DJs used vinyls, usually in a single format. The reason for using vinyls included better sound quality and availability (all of the DJs specialised in 1970's music). Their collections ranged from approximately 500 vinyl singles, to 10,000 vinyl singles. The number of genres also varied, with two of the three DJs specialising in one genre, and the third naming around 15 different genres.

2.3 Results

The results for each of the cases is presented below. Each case has a description of the seeking tasks, organisation methods, and strategies. Next, an overview of the main themes is given. The results were analysed using a Grounded Theory approach.

2.3.1 Case 1: Casual Users with Electronic Collections

Organisation Tasks and Strategies

Three systems of organisation were identified: self-organised, organised by an application, and not organised. Self-organised collections were generally organised alphabetically by either a topic property such as album or artist name, or by sound property, such as genre. Participants described how they created a file and folder structure to arrange their collections. Rules for deciding on what music would go where were decided upon by participants. This group of participants used other applications to play their music (e.g., Winamp [99] and Window's Media Player [100]). Some participants reported that maintaining a particular organisation was not possible, either because of the difficulty of organising some of the music, or a lack of time.

The second type of participants' music collections were collections that had been organised automatically by an application (e.g., iTunes [101]). To automatically organise music, tools often use pre-defined ID3 tags [102]. These tags include details such as the artist name, album name, number of tracks, lyrics, title of tracks, and genre. Problems that occur are when the ID3 tags are not accurate or missing. In this case, one participant reported having multiple tracks with the same name (e.g., track 1). Other ID3-based problems highlighted by the participants included misspelled names, genres that were too specialised, and album mix-ups where songs from one compilation album get distributed into several individual solo albums. Correcting these errors was seen, by some, as an onerous time-intensive task, and therefore not completed.

The third group used neither an application to organise, nor one to create a file/folder structure. All music went into one folder without any additional annotations (e.g., if a music file was automatically named "Track 1", it would remain that way). Usually, an alphabetical ordering of the file names was imposed by the operating system.

None of the participants reported duplicating the mp3 or creating short-cuts to help in organising their collection. No participants reported using automatic feature extraction tools to support them in organising their collections.

Seeking Tasks and Strategies

Typical tasks for participants were: to look for a specific track, artist, or album; browsing for music in order to create a playlist, to suit a mood or task, and finally; to browse for inspiration.

Common strategies for a specific re-find were to look where the object was expected to be, scan through a list of files, and enter keywords into a search engine to find a file. Frequent problems occurred when the text data attached to a file did not contain the details that the participants could recall. This meant that participants either browsed in the hope of seeing a file name that would prompt their memories, or listened to each track one after the other to find what they were looking for.

When creating a playlist, the audio properties were frequently the most important aspect for participants. Frequent situations that playlists were created for included, music for work, music for cleaning, music for parties, and music for relaxation. To generate playlists, a combination of re-finding and browsing was used. The second group of participants, who used tools to automatically organise their collection, frequently applied automatic playlist generators. These tools created playlists based on properties such as, most recently played, most often played, and randomly selected tracks (shuffle feature). Re-finding music to suit a mood or occasion was reported to be a more common task than finding a specific track.

2.3.2 Case 2: Casual Users with Physical Collections

Organisation Tasks and Strategies

In many cases, the physical organisation of a collection depended on the perceived type of use of the CD. The perceived type was divided into different sections: frequently used, CDs for on the move, less frequently, and never used. Frequently used CDs were placed physically close to CD players, often organised in piles. CDs for on the move were stored in a wallet, and were meant for playing away from the participants' home (e.g., in the office). Less frequently played CDs were kept in close proximity to a CD player, but stored in a CD tower or bookshelf in some order(e.g., alphabetical, order bought in, order listened to in, etc.). Lastly, CDs that were never listened to were stored out of sight. For example, one of the participants stored his seldom-listened-to heavy metal collection in the attic. Generally, the degree to which a collection was organised depended on its size, and on personal preferences. One of the participants tended to always put CDs back into the wrong case, and occasionally in one evening return them to their right cases. Visual cues from the CD cases gave clues as to what was on a CD, and was used for browsing purposes. If the collection was small, no organisation was created; for larger collections, genre or alphabetical ordering was a typical organisation technique.

Seeking Tasks and Strategies

Common seeking tasks were re-finding a specific CD, either to play or to lend to a friend and re-finding music that would suit a participants mood or for a particular occasion, for example, music for work or parties. When looking for a CD, participants reported looking first where they expected to find it. If it was not there, they would try to remember where they had last seen it. Once having found the general location, participants reported looking through a group of CDs until the desired CD was found. When not sure what music to listen to, looking quickly through the spines of the cases gave some participants inspiration. In the case of the participant who kept most of her collection in CD wallets and the rest in their individual cases, unless when looking for a specific CD, she often chose a CD that was in an individual case because it was faster to look through that part of her collection than to go through the CDs in the wallets. Most of these tasks were completed without listening to the music beforehand. Participants remembered different details concerning their music, such as the colour of the case, cracks in the case, how a track started, or lyrics. Listening to music was considered to be an enjoyable activity and in many cases, exact re-finding of a CD was not needed.

2.3.3 Case 3: Professional User with an Electronic Collection

Organisation Tasks and Strategies

Both the virtual instruments and sample collections were pre-organised by the musicians who had initially created the samples and virtual instruments. Their organisation started on very general terms, such as type of instrument, and worked towards minute details, for example, "a muted trumpet playing C in a large theatre". All samples were imported into a sampler for later use. Presets of commonly gathered sounds were also created. For example, to recreate a piano using samples, each key on a keyboard was matched to an equivalent sample. There was no need to look for each sample in turn, as all samples related to the virtual instrument were stored together. When composing, the composer used a palette of samples. The palette was created by listening to all of the samples in turn, and saving them in a folder on his desktop for future use. He described this task as a long and tedious that could take days to complete.

Seeking Tasks and Strategies

Common tasks were finding an exact match (e.g., "a sample that sounds like an old-fashioned telephone"), or browsing for samples to suit a particular idea. A linear search was the most common way of finding suitable samples. This was due to no automatic search facility existing (that he knew of), and that remembering all of the details of such a large collection was impossible. While writing, he frequently experimented with different types of sound, at times selecting random samples in order to experiment further. In this case, the lack of adequate search tools meant that the only option for finding new samples was to start at the beginning and work to the end.

2.3.4 Case 4: Professional User with Physical Collection

Organisation Tasks and Strategies

The extent to which the collections were organised depended on their size. One DJ described an elaborate system of storing his large (almost 10,000 vinyls) collection. Before organising, all vinyl sleeves were replaced with a white sleeve he made annotations on. Annotations consisted of relevant information such as the band name, album name, and memorable audio features of the music. Singles were organised in two places. Archived, less frequently played vinyls, were in alphabetical order placed on shelves. Frequently used vinyls were organised in one of three piles on the floor depending on their BPM (how fast they were). Records near the top of the pile were listened to more often than those near the bottom. Each pile consisted of around 200 records.

The other two DJs had smaller collections and a less systematic method of organising them. Both spoke of remembering visual cues, for example, a tear on the right hand corner of the sleeve or the last location a vinyl single had been played. Both organised their collections less frequently and arranged the vinyls alphabetically. All of the DJs took around 200 vinyls to work in a bag.

Seeking Tasks and Strategies

When working, common tasks for re-finding were specific requests from the audience, browsing for music to suit the mood of the audience, and looking for a specific vinyl. For both of the specific re-find tasks, all DJs spoke of having an excellent memory. They knew if they had a track and where it was. For the browsing task, they had time (a few minutes) to go through their bags to see what they had. Again, all said that having an excellent memory helped them remember various features of the music. All vinyls used were singles, so they did not need to remember if the music was on a compilation vinyl or not. At home, all DJs stated that re-finding parts of their collection was not a problem, as either they would not look for a specific item, or their excellent memories aided them in finding any relevant singles. Instead, all of the DJs highlighted the importance of being able to find new music in public collections.

2.4 Discussion

The main findings of this study were: passive information seeking was more common than active; physical collection organisations were frequently organised spatially, and evolved over time, whereas electronic collections were organised using classifications. Three groups of user became apparent: manual organiser, automatic organiser, and non-organiser. Finally, no automatic feature extraction tools were used by the users for electronic collections.

2.4.1 Organisation

The results show a number of behaviours exist when organising music collections. Whether a collection was organised and maintained or not, appears to have depended on the perceived cost of organising. If the perceived cost in time and effort was considered to be too high, then the collection was not formally organised, but had an emergent organisation instead. An example of a collection where the cost of organising it may be too high is a relatively small collection that its creator was familiar with. Emergent organisations were typically organised informally by the frequency tracks are listened to or when they were purchased. Frequently, organisations were not maintained. This was noticed with the participants who put CDs in the wrong place, and with the participants who did not re-name ill-named mp3 files. Automatic features, such as automatic extraction organisation, were not applied by the users. Neither was copying music in order to put it in more than one place.

Three of the cases highlight personal music collections as consisting of different types of information that is organised depending on how often it is used/listened to. For example, music that is listened to often, sometimes, and infrequently. This was noted strongly in the case of physical collections, but less so in electronic collections. The participant that described using different information types, was the composer. He described his audio samples as being categorised as frequently used, working, and archived types. The frequently used samples were the samples he was currently using. Working samples were samples he had used previously and still used regularly. Archived samples were samples that were infrequently used. This demonstrates that information types that exist for paper and electronic documents also apply to music collections, and it indicates that spoken word collections may also be organised depending on similar information types.

Physical collections provided a richer set of organisation methods, including classification, emerging structures and spatial organisation. Classifications used in the physical collections consisted of alphabetical or genre ordering. Spatial organisation included CDs being kept where they were most likely to be played (e.g., at work, in the car, close to a CD player). Less listened to CDs were archived on a different shelf. In the case of professionals, music for work was kept separate from music for personal entertainment. Spatial placement depending on use was also noted by Cunningham et al. [83]. In physical collections, spatial organisations, similar to emergent organisations, may emerge over time because of the physical nature of a CD. In electronic collections, this organisation is presently less obvious. An example is the use of the desktop for storing frequently used information [75]. This was reported by the composer. The range of the described strategies shows that individuals behave differently. They develop strategies over time that suit their needs best.

2.4.2 Seeking Tasks

Re-finding music to suit a mood or occasion was a more frequent task that that of re-finding a specific CD or mp3. All cases described a problem of re-finding music where the main details were not known (e.g., participants remembered how the chorus began or if it had a guitar solo, but not the name of the track). Participants reported remembering different details concerning the music, from the cover of a physical CD, to some of the lyrics. Often, for casual users, an exact fit was not essential. For the composer it was more important.

Tasks that were common in both types of collection were specific known-item searches, best-fit searches, and groups of best fit searches. The most frequently used modes were exploratory or browsing. Often, finding the right object was not as essential. Evidence suggests that specific finding tasks are less frequent than browsing tasks. Studies by Laplante et al. [88] and Lee et al. [42] showed that users seeking music used a browsing strategy. They showed that music seeking was a non-goal and enjoyable activity, with most of the tasks being completed being a browsing task.

Empirical research has not yet identified how best to support navigation within collections, although a number of techniques have been suggested. Navigation is inherently a spatial metaphor [103], and therefore many of the techniques proposed to aid navigation have physical-world counterparts. Wayfinding techniques aim to aid users in finding their way within an information collection. In order to successfully navigate within an information space, people need to remember where they have been, as well as where there may be a possible interesting route.

2.4.3 Importance of Memory

Browsing requires a good memory to know where you have been and were you may go. In the case of re-finding, memory was indicated to play an important role. Another point was that users tend to remember more general details about information, rather than specific content. Memory cues are stronger in physical collections due to the additional number of attributes an object may have. In electronic collections, users are required to remember more details to find what they are looking for. There are a number of ways that support memory: how information is organised, how well file are named, noting visual features, and location. The implications are that it is important for tools to leverage skills that users already possess.

Memory is proposed as an important success factor in re-finding relevant information. Elsweiler et al. [91] proposed that lapses in memory impede users from successfully re-finding information. It has been noted that people remember the general idea within a piece of information, rather than the specific details [104, 105, 106]. This means that, the query-based approach is less successful than a location-based approach because users memories do not work by recalling known details of information, but call up an impression of the information and context [97]. Memory may fail because recollection diminishes over time (known as decay theory [107]), and if users are focussing on other tasks and interacting with other objects (known as interference theory [108]).

Different types of memory have been identified by Elsweiler et al [44]: semantic memory, autobiographical memory and temporal reference. Semantic memory is when people recollect properties of information, such as its name. Autobiographical memory is when information is remembered as a previous experience. Temporal reference allows people to remember when an object was last seen. Elsweiler et al. argue that PIM tools should support these different types of memory.

Few of the cases presented here described difficulties in finding particular music within their collections. Cunningham et al. [83] in their music collections study also noted that users did not report of having many difficulties locating what they wanted. Usually they could find it within a few steps.

2.5 Chapter Summary

This chapter described seeking tasks and strategies that are expected to be relevant to spoken word collections. Collection organisation was highlighted as an important factor that would affect seeking strategies. A number of styles and strategies of organising a collection were highlighted, and it is expected that the way in which seeking tasks may be supported are dependent on how a collection is organised. Questions that still need to be answered with regards to spoken word collections, include how much longer it takes to perform seeking tasks; how users decide if an object is relevant; what types of information is remembered; and is playback of an object necessary. The next chapter provides an overview of tools and concepts that aim to support the seeking tasks and strategies identified in this chapter.

Chapter 3

Supporting Seeking Behaviour in Collections

Chapter 2 described seeking tasks and strategies commonly used when finding and re-finding information within text and music collections as well as providing an overview of collection organisation and the styles and strategies used (or in some cases not used) within the organisation of personal collections. The main findings were the differences in user preference when organising and maintaining personal collections, the use of spatial and emergent organisations within physical collections, and classifications within electronic collections. The most common seeking task within personal music collections were re-finding by browsing where an acceptable match was sufficient. Seeking specific information was also highlighted, but was not performed as frequently. Supporting user memory and providing quick and easy access to audio content was also highlighted as important factors. This chapter gives an overview of tools and concepts that aim to support seeking tasks and strategies within information collections. With spatial organisation highlighted as a desirable method to organise physical collections, there is particular emphasis on spatial organisation and visualisations tools that support audio collections when possible. The tools described here are not meant as an exhaustive list, but merely an illustration of concepts. Different features aimed at supporting natural capabilities are also discussed in this chapter. A comparison of audio and visual information collections is also given, with possible uses of audio as an information carrier highlighted.

This chapter is organised as follows: Section 3.1 describes tools and concepts that support the three organisation styles identified in Chapter 2 (*manual organiser*, *automatic organiser*, and *non-organiser*), providing a list of possible features that may be included in a spoken word collection tool. It also describes how these tools support different navigation and seeking tasks discussed in Chapter 2. Section 3.2 first describes the main differences between audio and visual information. It then describes non-visual visualisations (collections where information is presented using only the audio medium). Finally, it presents an overview of concepts developed within audio research aimed at transferring information in audio as opposed to through visual channels.

3.1 Information Seeking Tools

This section introduces the organisation strategy types, defined in Chapter 2 (i.e., *manual organisation*, *automatic organisation*, and *no organisation*). For each case, it describes how the common navigation and seeking strategies are supported within these types of tools.

3.1.1 *Manual Organiser*

Manual organiser is the strategy of organising collections without the help of automatic tools. Typically manually organised spatial organisation tools provide a workspace for users to store and organise their information. The act of organising the information can improve the understanding of how the information within a collection relates to each other. The spatial structure can be used to hold collections of information that are roughly structured, and acts as a reminder of tasks that need to be completed. Tasks that are well-suited for manual spatial organisation tools include those that have a difficult-to-define goal, such as knowledge building and decision support. In most cases information collections are smaller and are manipulated by fewer users. Manually organised spatial organisation tools are particularly useful in supporting the use of ephemeral and working type information. Less is known about how effective they are for archived information. This makes tools that support manual organisation favourable for knowledge work where analyses or decision-making tasks are an important part of the task. One reason for this is the ability of manually organised spatial organisation tools to support emergent problem solving [67]. It is reasonable to assume that knowledge work type tasks that use spoken word collections will be equally well supported using manually organised spatial tools. Many of these types of tools are based on a card-on-the-table metaphor, and provide additional visual cues to help users remember connections between information.

Notecards [109] is an early type of manually organised spatial tool. The creators noticed that users organised information spatially when organising in the physical world. It was designed to support emerging structures and, therefore, for tasks that were ill-defined (see Fig. 3.1).

Later, VIKI [110] introduced a fisheye lens into its spatial organisation to help users in exploring large information collections. With the lens information in the centre appeared larger than that on the outside. This meant that more information could be placed within a user's field of vision (see Fig. 3.2).

The Visual Knowledge Builder(VKB) [68] extended the number of visual cues enabling users to organise not only by position and proximity, but also by colour, size, and shape (see Fig. 3.3). The VKB also allowed different types of information to be organised, for example text, webpages, pictures, and (most recently) audio files.

The use of manually organised spatial tools are being investigated within personal spaces in digital libraries. MiBiblio [111] and Garnet [112] are examples of spatial tools

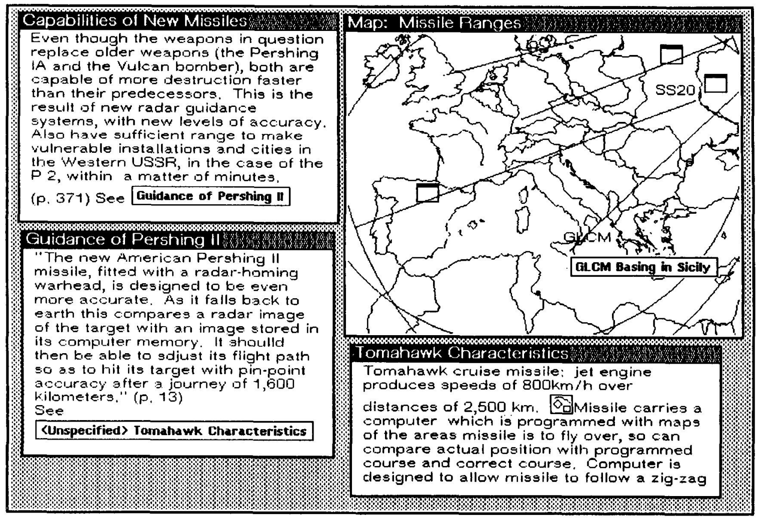


Figure 3.1: User interface from Notecards displaying four related information objects.

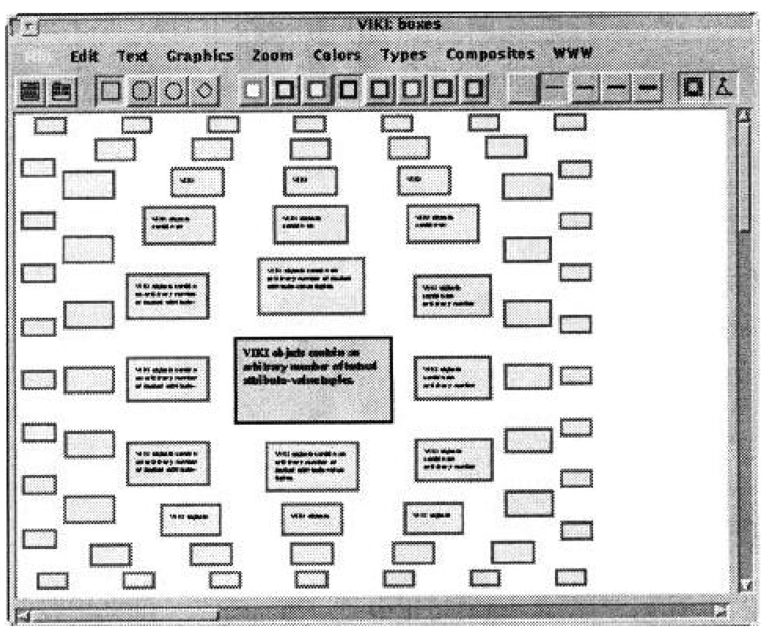


Figure 3.2: User interface for VIKI showing a multi-focus fisheye lens.

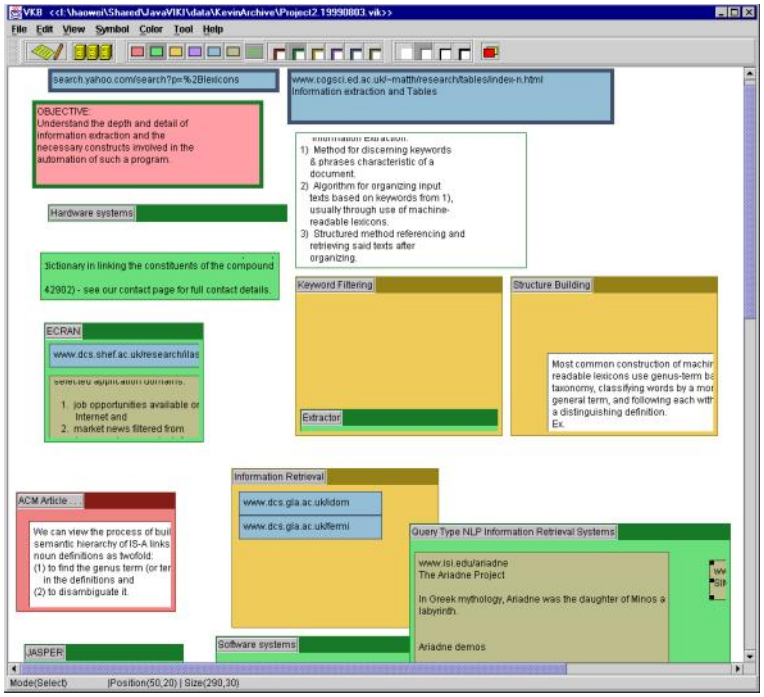


Figure 3.3: User interface for Visual Knowledge Builder (VKB). Different visual attributes show possible relations between the shown information.

used in digital libraries (see Figs. 3.4 and 3.5). MiBiblio allows its users to create personal spaces with services that are important to them. These services are represented as books on shelves. It is also possible to include articles, reserved books, and notes within the personal space. Garnet supports the personal organisation of materials within a space. Search features are supported by being able to save search results and group relevant items together.

A difficulty of spatial organisations is that as the size of a collection increases, users may experience visual overload. As a consequence the time it takes to find/re-find wanted information may increase, the error rate of successfully finding/re-finding may increase, and the level of satisfaction at using such a tool may decrease. One method of reducing the difficulties of information overload is to introduce additional visual cues in order to provide a large distinction between information groups, such as fisheye lenses, and focus and context (described in Sect. 3.1.2).

Figure 3.4: User interface for MiBiblio. Books represent available services for users.

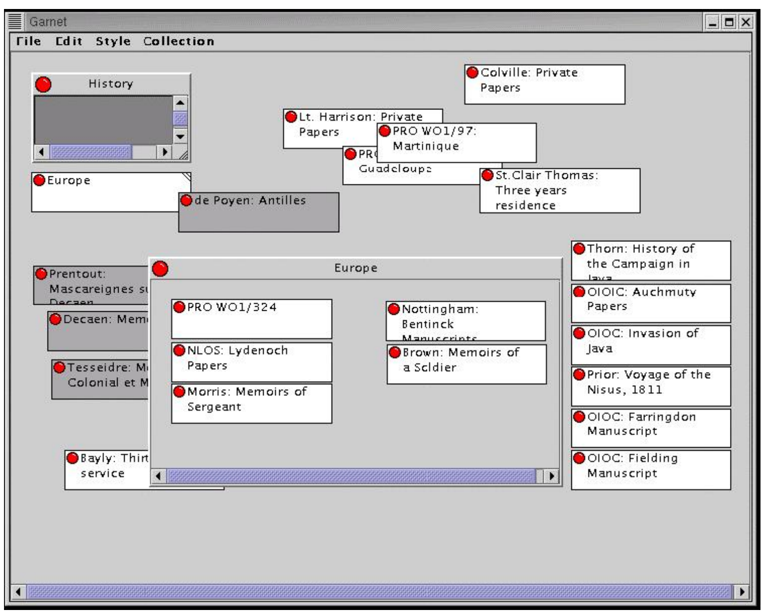


Figure 3.5: User interface for Garnet showing a saved search.

Another difficulty generated by visual overload is not being able to remember, recognise, or represent all of the relationships between separate information objects. One way of increasing the number of found relationships is to introduce a tool that suggests possible connections between information, based on how the information has been organised and by adding visual cues. The VKB includes such a tool that analyses users' organisation patterns and suggests possible relationships. It also includes a history function, where users may move back and forward in their time line in order to improve their analyses.

Another technique used to reduce the difficulty of visual overload is to increase the number of dimensions available. Data Mountain [113] is an example of a 2.5D spatial organisation tool (see Fig. 3.6). Users move miniature representations of webpages in a 2.5D space. The further away the icon is moved, the smaller it becomes. Studies showed that users were faster at re-finding relevant webpages within this spatial workspace, than when using the traditional lists provided in bookmarking tools [113]. Additional studies using Data Mountain have shown using 2.5D version is an improvement on 2D spaces. Three-D spaces appear to be less effective [114] than 2.5D spaces. Cockburn et al. suggested that as the dimensions increased, the additional complexity confused users [115].

Until recently audio information has not been supported within these set of tools. However, they way participants described how they organised their physical collections in Chapter 2, implies that these spatial tools would be equally relevant for them. In order to make these tools more useful, easy access to the audio content is required. Possible approaches are discussed in Section 3.2.

Navigation strategies described in Chapter 2 can be described in terms of breadth and depth, and teleporting and orienteering behaviour. At present, most research within information collection navigation focuses on the Web, and therefore on collections that are organised either by association or by classification. Within web research, a navigation

Figure 3.6: User interface for Data Mountain showing a collection of webpages organised manually in a way that reflects a users preferences.

system can be described as either being embedded or supplemental. An embedded navigation system can either be at the global, local or context level. Whereas a supplemental navigation system provides different ways of accessing the same information. Examples include, sitemaps, indexes, and guides.

Examples of navigation systems used within manual organisation tools include the use of a history function allowing users to move back and forward in time, (similar to the back and forward buttons on the browser). However, much of the emphasis of these tools is on memory support, where it is assumed that the time taken to organise a collection strengthens memory cues allowing easier access to the "right information".

Typically, seeking tasks within manual organisation spatial tools assume that users are acquainted well enough with their collections and have a good mental model of the space. The most common tasks are browsing or searching re-finding where the level of fit (i.e., best fit to acceptable) and amount of desired information varied. For the tools described here, it is proposed that the time taken to organise a collection means that users' memories concerning the location of their files improve and, therefore, users should be faster at locating wanted and known information. Browsing tasks are supported by relevant and related objects being placed together. Frequently, search engines are not available for these types of tools.

3.1.2 *Automatic Organiser*

Automatic organiser is the strategy of employing tools that automatically organise collections either with or without well-defined guidelines. Spatial organisation tools that automatically organise collections are frequently known as visualisation tools and organise a collection based on pre-defined, well-understood rules. These tools may either be used for personal or public information collections.

An important principle for most visualisation tools is “Overview first, zoom and filter, details on demand” [84]. The overview is intended to give users a comprehensive view of the entire collection and help navigating around it. A number of strategies for doing this include: the use of maps (see Fig. 3.8), fisheye lenses [116] (see Fig. 3.7), and focus plus context [117] (see Fig. 3.9). A multi-modal way of providing focus plus context is implemented in Dolphin [118]. Visual cues provide focus, and the context is presented using audio cues, taking advantage of the natural affordances of both audio and visual cues. Zoom refers to allowing users to move in towards items of interest. Filter supports users in reducing the number of visible items that are not of interest. Strategies include the use of sliders, buttons, and other control widgets combined with a rapid display update. Often, a query is combined with filtering. Finally, users are presented with the details on an individual node when wanted. Other tasks that are frequently supported are: relate, history, and extract. Relate shows the relationships between individual items. History supports the undo and redo functions and progressive refinement. Extract supports detailed searching within a subset of the collection. Usually, there is an ability to save search information to a file.

Figure 3.7: Example of a fisheye lens.

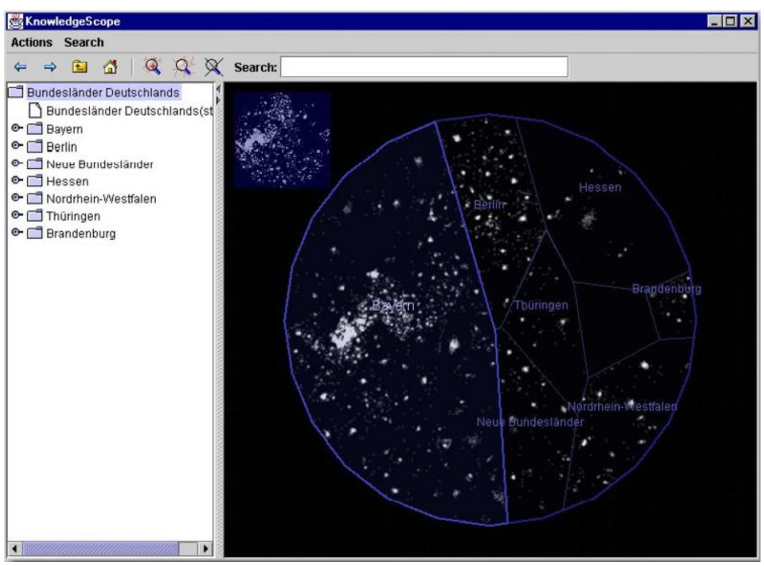


Figure 3.8: User interface for InfoSky showing the whole space, as well as a map in the corner that gives an overview of the whole space.

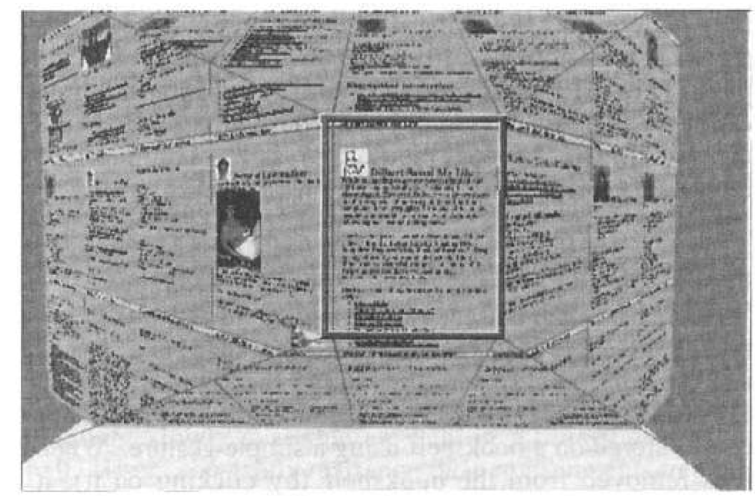


Figure 3.9: Example of focus plus context in the Web Forager.

There are a number of methods for visualising information in order to provide an optimal visualisation, Shneiderman recommends deciding on what type of information needs to be visualised and matching the type to a well-known visualisation technique [84] (see Sect. 2.1.2).

For this work, a spoken word collection can be described as a network. Even though an individual spoken word document would be classified as a temporal type, the focus of this work is on representing collections, and not the internal spoken word document structure. A number of techniques have been investigated to support network information, such as visualisations suited for trees (providing the network is reduced to a tree), force-directed placement, and energy-based placement. Many of these techniques involve clustering. Clustering techniques organise collections in such a way that groups (also known as clusters) form and highlight how pieces of information relate to each other. The main idea is to group related objects together, while at the same time keeping unrelated objects further apart. There are many methods to create clusters, (see [119] for a review). Islands of Music [120] (see Fig. 3.10), PocketSomPlayer, and PlaySom [121] (see Figs. 3.11) use self-organising maps (SOMs) [122] to organise a collection by genre using feature extraction. Similar genres are positioned closer together in order to improve seeking by

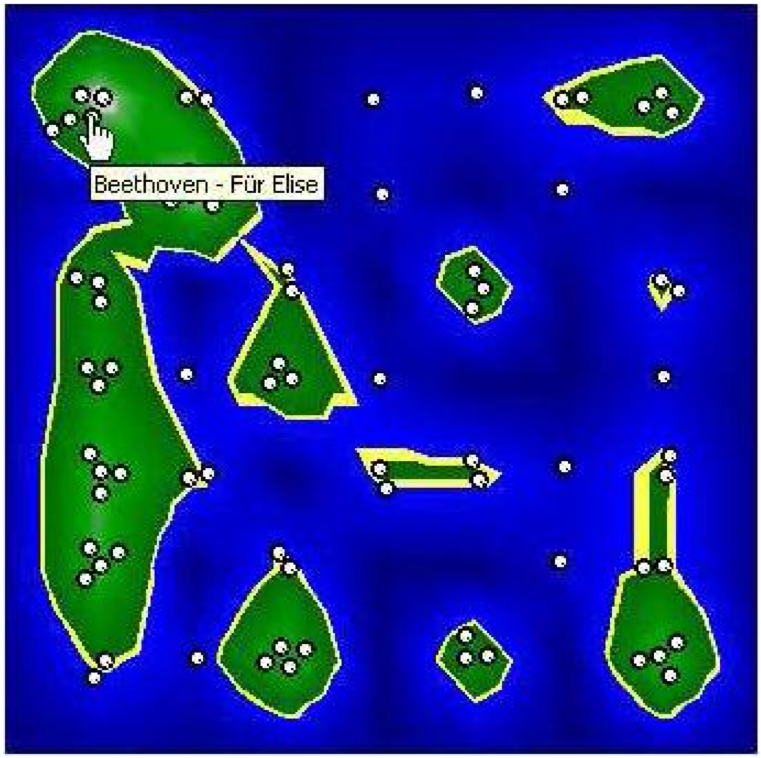


Figure 3.10: User interface for Islands of Music.

browsing. Users can also estimate how large each genre is, and potentially how long it would take to explore.

Tree Visualisation Techniques

A network is essentially a graph. Using known methods, graphs may be reduced to trees, and visualisations designed for trees can be used. Examples of tree visualisations are treebrowsers, treemaps, radial approaches, cone trees, landscapes, and hyperbolic browsers (see [123] for a review).

The Audio Retrieval Browser (ARB) [124] supports users in browsing music collections using tree visualisations. Individual music objects are represented as both visual and audio objects. These objects are displayed in a browsing space. Users can map different properties of the music, such as artist and genre, to visual cues, such as colour and shape.

Figure 3.11: User interface for Playsom and PocketPlaysom. Example of SOM in music clustering. The image on left shows the desktop version, the one on right the mobile version.

The information can be viewed in either a 2D or a 3D soundscape that provide interactive multiple audio playback (see Figs. 3.12 and 3.13).

Force-directed placement

Force-directed placement is based on physical properties of information and consists of two components: a model and an algorithm. The model consists of physical objects and the interactions between them, and the algorithm computes the equilibrium for the system. One method of creating a force-directed placement is the spring model. Nodes in a graph are joined using imaginary springs that gradually move the collection into clusters. An example of a force-directed placement is shown in Fig. 3.14.

Energy based placement

Energy based placements build on the idea of spring models. Seeing Sounds [125] is an example of a energy based placement. It supports users in exploring a pre-organised

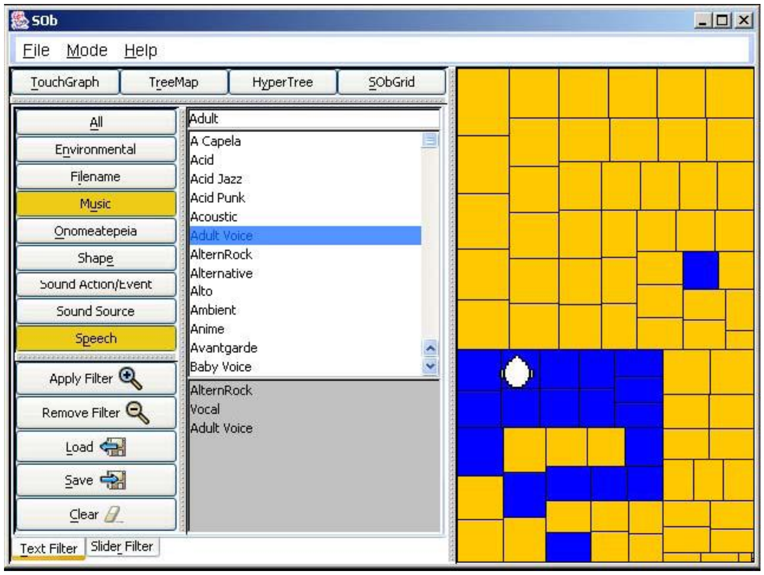


Figure 3.12: User interface for ARB showing a TreeMap visualisation.

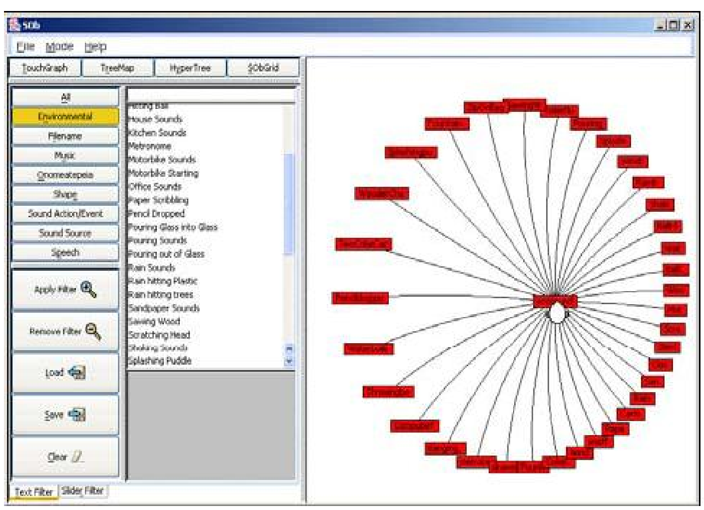


Figure 3.13: User interface for ARB showing a HyperTree visualisation.

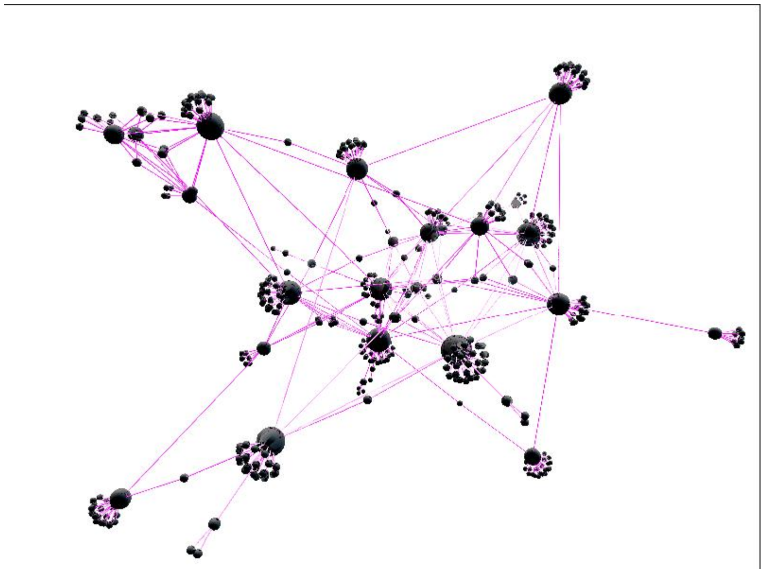


Figure 3.14: Hyperspace is an example of force-directed placement visualisation.

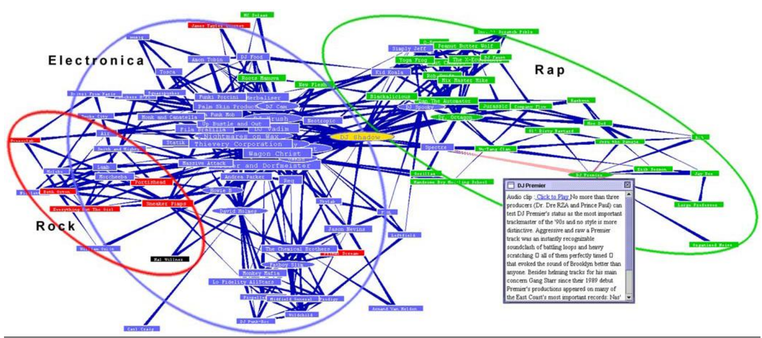


Figure 3.15: User interface for Seeing Sounds.

music collections network (see Fig. 3.15). The aim of Seeing Sounds is to improve the browsing and search facilities of music collections by clustering related music. Users hear short clips of the music by opening an additional application, and each music node gives users some biographical information about the music. Instead of text representative tracks are chosen to best describe the artist. Users can listen to more than one music excerpt simultaneously.

Sonic Browser [126], similar to Seeing Sounds, supports users in browsing music collections. Users listen to sound excerpts by hovering with the mouse over the node. Users

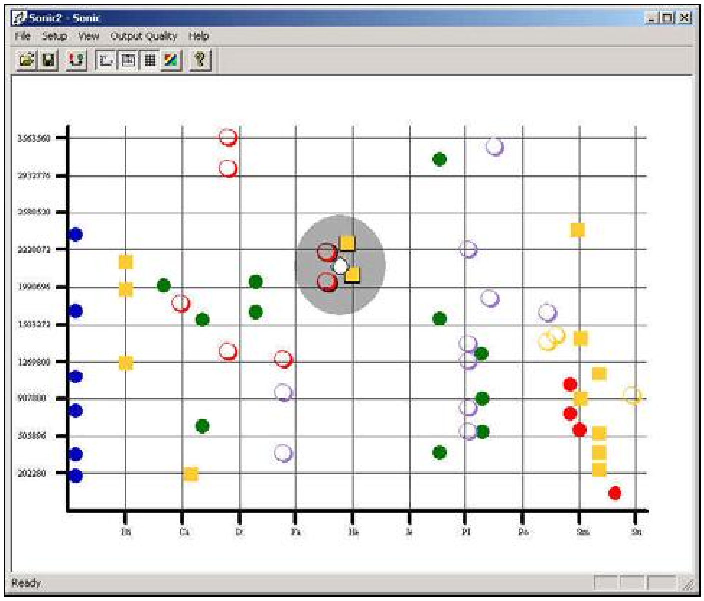


Figure 3.16: User interface for Sonic Browser.

are able to map sound features to different visual attributes, so they can see what sounds are similar (see Fig. 3.16).

Within visualisation tools there is more scope for supporting navigation. By using these tools it is possible to organise a large collection automatically; therefore, the memory cues gained through manual organisation are not present. This implies that the navigation cues offered must compensate. As previously mentioned, many visualisation tools are based around the concept of providing overviews and various navigation features, such as maps, zooming functionality, extraction, and history functions.

Visualisation tools support both searching and browsing tasks by providing both query searches, and grouping related information together. These collections are also being discussed for music collections but have so far not been implemented for spoken word collections. These collections are expected to be useful for both personal and public spoken word collections.

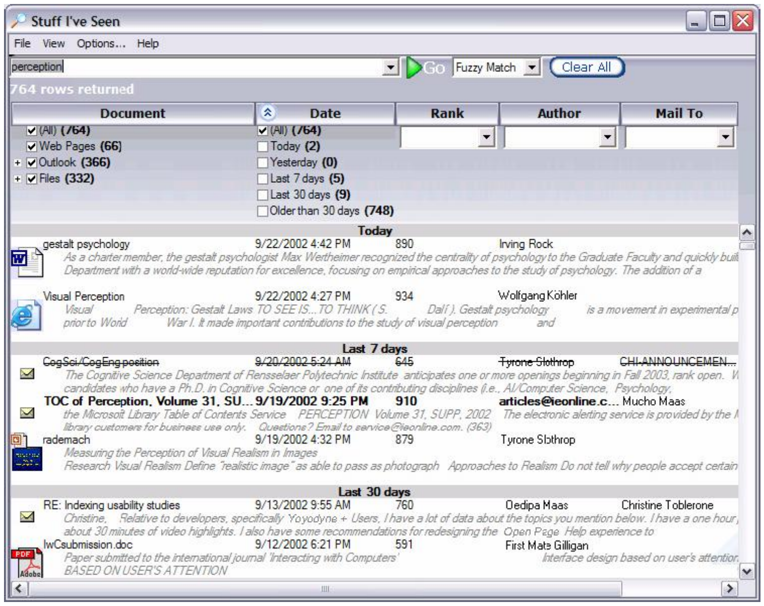


Figure 3.17: User interface for Stuff I've Seen (SIS) showing the different searching possibilities and results from a search.

3.1.3 *Non-organiser*

The third identified user group type was the *non-organiser* type. This type of user is characterised by their lack of organisation, either manually or automatically, preferring to use their time to retrieve wanted information when needed instead of organising it in some way first. This technique is used widely in both personal and public collections, and is based on the assumption that it takes less time to look for the relevant information at the point when it is needed than to organise a collection first. Users who do not organise their information rely on such tools to find their information quickly. Recently, a project called "keeping found things found" [127] has investigated seeking using query-based tools instead of organising information. Stuff I've Seen (SIS) [128](see Fig. 3.17) and the Haystack Project [129] (see Fig. 3.18) are tools that support this idea. Other tools that allow efficient searching instead of organisation are desktop search tools, such as Google desktop and Filehawk (see Fig. 3.19).

These tools work well for search tasks when the "right" information about an object

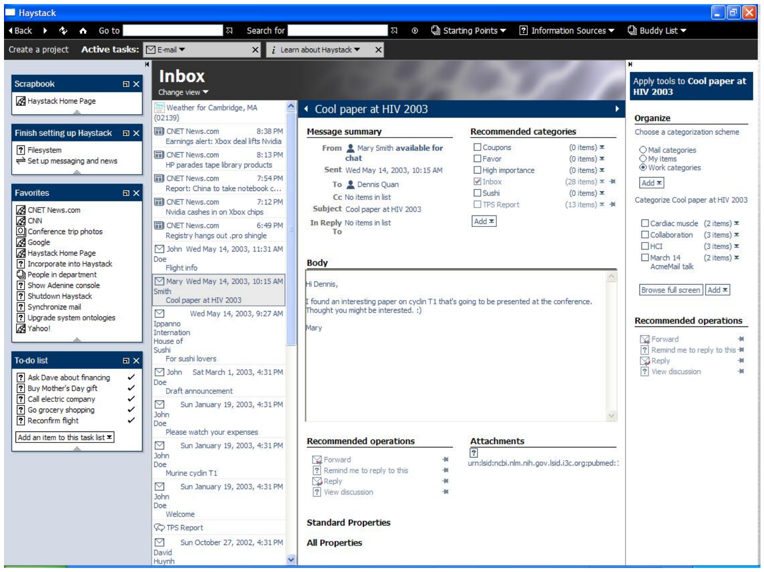


Figure 3.18: User interface for Haystack showing the e-mail window and searching possibilities.

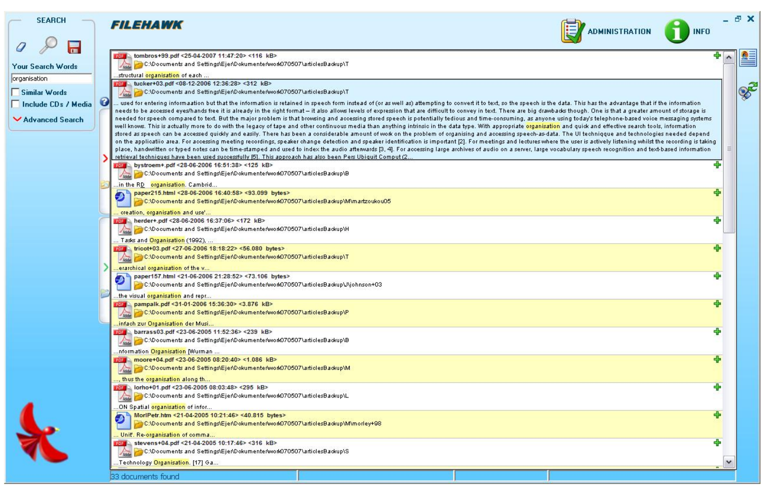


Figure 3.19: User interface for Filehawk showing the opening search page, a query and returned results.

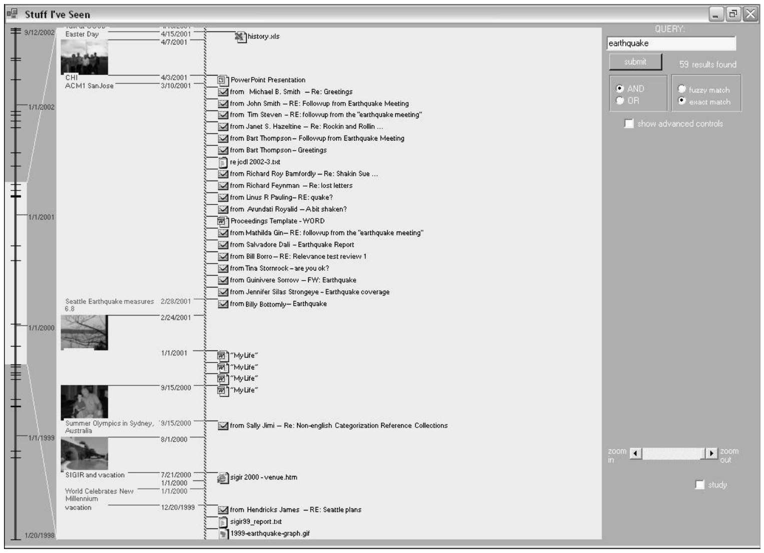


Figure 3.20: Stuff I've Seen (SIS) displaying landmarks within the interface.

is known, but may be less efficient in supporting browsing tasks which typically rely on the placement of related information in order to encourage exploration and support memory. A method that SIS uses to support memory is that of Landmarks. Similar to landmarks in the physical world offering a reference point, electronic landmarks aim to provide reference points for information. SIS provides both temporal and spatial landmarks. These landmarks are created by its users on personal information. The landmark can consist of any type of information and act as a memory cues to users. A temporal landmark references a specific point in time; for example, a picture of a holiday 10 years ago is a temporal landmark. Spatial landmarks reference placement. In a spatial tool, a spatial landmark could be a particular cluster with a distinct appearance. A study performed using SIS showed that participants retrieved information more quickly when using a combination of landmarks and dates, than dates alone.

Clusty [130] is an example of a search engine designed to support browsing (also known as exploratory search) rather than searching tasks. After entering a query it presents the results along with relevant categories (see Fig. 3.21). Research in the area of exploratory searching is relatively new, with workshops such as HCIL SOH '05, SIGIR'06,

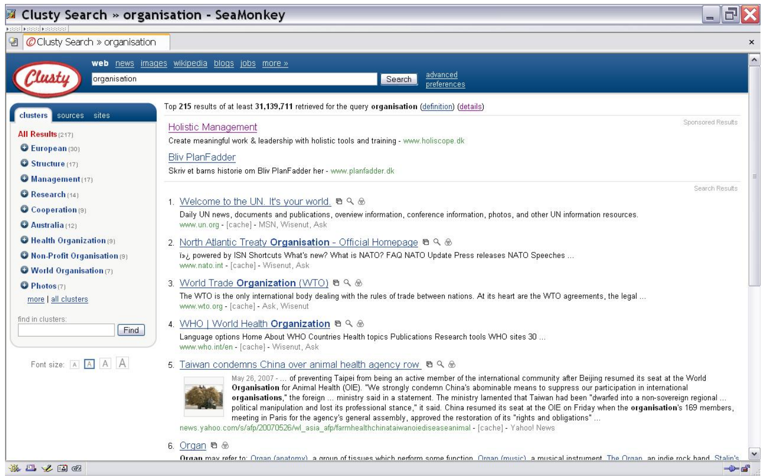


Figure 3.21: User interface for Clusty.

and CHI'07 being set up to better define what exploratory search is and how to design tools to support it. As of yet, metrics for testing tools against exploratory search do not exist.

Many music websites are designed to be browsed by their viewers. Typical techniques are to group related music together (similar to the ideas employed by visualisation). Different groupings are apparent in these websites, including classification by genre, band, artist, new releases, favourites, and recommended groups (e.g., iTunes see Tab. 3.22).

There has been much research on text-based information, but when searching for spoken word documents, using search engines may be more difficult. If users can search using text-based meta-data, spoken word documents may be retrieved using a standard query-based search engine. Alternatively, automatic speech recognition (ASR) tools can transfer enough of the audio into text for it to search using a query system. Tools such as PodZinger [41] and TVEyes [40] use ASR to search the audio content before returning the original audio file. SpeechBot [131] is another example of an audio search engine using speech recognition. It searches over 17,000 hours of audio content from websites and gives users links to the audio content containing their search. At present, such tools are available for public collections, but not for personal information.

Figure 3.22: Opening website for iTunes.

Another problem with these search tools is that although they work well for re-finding tasks that are well-defined, they are less effective when users' cannot efficiently articulate what they are trying to re-find, or when they do not know the correct details, for example, "re-find the file I was looking at last week that had a diagram showing seeking tasks", or worse, "the music I was listening to that had a particular rhythm". Content searches within text are becoming easier to perform, but at present it is not possible to do the same for spoken word.

3.2 Use of Audio within Spatial Tools

A difficulty with using visualisation or spatial representations, is that as the amount of information grows, and the complexity of the information collections increase, users may become visually overloaded. Visual overload increases the time spent and error rate in locating information. It has been suggested that complex visualisations can be created, but users inability to process these complex visualisations renders them less helpful. Approaches to reduce this, include adding more dimensions (as described in applications such as Data Mountain), increasing the number of visual and navigation cues, and transferring part of the information to a different modality, such as audio. This section focuses on transferring part of the information to audio. Naturally, with spoken word collections, it is highly desirable that the contents can be listened to easily and quickly. The section describes ways in which audio may be combined in a visualisation to improve user performance.

3.2.1 Differences between Audio and Visual Modalities

Audio and visual information are different modalities and work distinctly differently from each other. In order to understand how and what to transfer to the audio domain, it is necessary to understand how each modality is processed, and that the advantages of each are. Visual information provides intricate detail but in a smaller area, whereas audio information tells your eyes where to look. Visual information is persistent, Whereas, if sound is stopped, there is silence. One possible combination is to use audio cues to provide an overview of the information, and visual cues for intricate detail (see focus and context use within Dolphin [118]). It is possible to listen to more than one sound simultaneously, but not to read more than one document at once. Studies in providing simultaneous speech have shown that users are capable of following more than one voice simultaneously providing the difference in voice type is clear enough and not too many speakers are talking at the same time [132]. This implies that when the information collection consists of audio information, presenting more than one information object simultaneously may reduce the time it takes to find the "right" information.

3.2.2 Non-Visual Visualisations

Non-visual visualisation maps information collections into an audio space using audio attributes. Metaphors for non-visual visualisation include bubbles [133] and hallways [134]. These metaphors assume that users are able to imagine these areas without necessarily seeing them. A difficulty of these tools is that audio is a temporal information type. This means that users need to be able to hear the audio information as often as necessary. In order for audio metaphors to be successful, audio cues need to be as transparent to use as visual cues, and offer as much useful information as their visual counterparts. Key issues in audio visualisations include control over the presentation of auditory information [135], finding and selecting links within audio information [136], and over-viewing the information. At present, most research that discusses the use of transferring information to the audio domain is created by researchers that focus on the visually impaired and users on-the-move communities (i.e., mobile and ubiquitous computer fields).

An early attempt at visualising spoken word collections was that of Hyperspeech [137]. It was an audio-only "visualisation" that provided a network of spoken word documents. Navigation was controlled by spoken commands. Later, the AudioWeb [136] provided a networked visualisation of audio information. In this case, the spoken word document was treated as a first class object, but users navigated the collection using different visual browsers (e.g., a phone menu type browser and a ticker type browser).

Another example of an audio-only visualisation is the Audio Hallway [134]. The visualisation for Audio Hallway is that of a hall that has rooms containing spoken word collections. As users navigate along the hall, streams of audio can be heard coming from each room. Each room describes a different category, and the combined sounds give an indication of the topic and emotional content of the collection. When a room has been selected, the spoken word collection is separated from the braided audio into its individual parts and played in an arc around the user. In this way, the Audio Hallway provides two "views" of the data: an overview and a more detailed view. The braided audio created as part of the Audio Hallway takes advantage of users' abilities to listen to multiple audio streams simultaneously. Espace [133] is a hierarchy-based non-visual visualisation. Each folder has an ambient sound attached to it. When users move into folders, they can hear which folder they are in. If they move within a folder that is already inside a folder, both sounds of the folders are combined.

3.2.3 Representing Information in Audio

There are a number of occasions where audio cues may be used, such as for feedback, previews, and to reduce visual overload by representing multi-variate data. Using audio cues for feedback is frequently used within personal computing (e.g., error message sounds, signalling when a task is complete), and to augment computer games, and virtual environments. Another use of audio is for previews, particularly in the case of audio information. In a music seeking study by Laplante et al. [88], being able to quickly preview music was highlighted as an important feature for potential music retrieval systems. Audio can be useful when the information belongs to multi-variate, multi-dimensional, complex or large collections [138].

Research within audio interface design has produced concepts such as multi-modal focus and context [139, 118], earcons [140], auditory icons [141], and sonification. Auditory icons were first described by Gaver in 1986. The basic premise behind them was to map everyday sounds to a computer event by an analogy (e.g., the sound of paper being thrown away when the recycle bin is emptied in Microsoft Windows). Auditory icons try to mimic real world sounds. This can reduce the learning period, but may have restrictions on how each object should be mapped to a sound. Some objects and actions provide a natural mapping, whereas others are less obvious. For example, what would a "redo" action sound like?

Earcons were first discussed by Blattner et al. in 1989. Earcons are "nonverbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation, or interaction" [140]. It is a language built of short sequences of tones and associated with actions and objects. Earcons give each object and action a short musical sequence, or module. For example, a folder will have a particular module, as will the action of opening it. When the two are combined it means that a file is being opened. The design of earcons is related to their visual counterpart, icons. An advantage of earcons is that the number is limitless. Difficulties with earcons include that they are arbitrary and therefore, tend to have a longer initial training period. Receiving all the information from an earcon is also slower than that of an icon. Possible methods of reducing the duration of earcon sounds is to present earcons simultaneously.

Sonification, or auralisation, is when data is transformed by a sound generator into classical dimensions of sound, such as frequency, amplitude, and duration, for monitoring and comprehension. Sonification gives users immediate information about their actions. For example, if a sound was mapped to downloading a file, the sound may be continuous while the file is downloading, but when the process is finished the sound would change. This tells users that the action is complete. This type of technique immerses users more into their environment and can illustrate the passing of time. However, it can lead to artificial sounds, as there may not be real connections between the sounds chosen and the data that is being represented. This is presently being discussed within areas of business analysis and programming.

3.3 Chapter Summary

This chapter provided an overview of spatial organisation and visualisation tools that are being used to support seeking tasks for the three types of user identified in Chapter 2. In general, designing for unknown, impossible to articulate, exploratory information seeking tasks presents difficulties. Browsing tasks are also more frequent than that of searching; however, known-item seeking tasks frequently have better support through the use of query-based search engines. In order to support exploratory seeking better, one approach is to group related items together, and make users aware of additional related resources. Another important feature is allowing users to preview the audio information quickly. The findings of this chapter suggest using audio cues in order to provide a quick advanced preview and presenting simultaneous audio as a way of reducing visual overload. Features that can be used within visualisation and spatial organisation tools that support memory include maps, panning, zooming, and landmarks in order to provide a good overview of the information.

Chapter 4

Organisation and Re-find Experiment

Chapter 2 identified three groups of users based on organisation choice. It implied that how users sought information depended on how their collection was organised and/or maintained and what type of information was in their collection. A number of seeking tasks were proposed within the general categories of browsing and searching, and different strategies were described by users when completing both tasks. Chapter 3 described tools that supported seeking behaviour within collections and attempted to highlight what features would be the most useful for the three different groups of users. It was recommended that spatial organisations that supported fast access to audio would improve the accessibility of spoken word collections.

This chapter describes an experiment that aims to replicate organisation and re-finding behaviour described by *manual organiser* users. Using the analysis of tools in Chapter 3, Personal Audio Work Space (PAWS) was designed. Features include two views of the information collection, landmarks, quick audio preview, simultaneous presentation of audio, and the use of audio to describe meta-data. The results of this experiment are presented in this chapter.

4.1 Motivation & Background

Besides users who have personal music collections, another type of user that uses a manual organisation approach is knowledge workers (e.g., oral historians, anthropologists, usability specialists, and market researchers). An advantage of organising manually is that the process of organising a collection can help to analyse the data better. With the growing collections of spoken word content, knowing how people organise and re-find spoken word documents and supporting these actions may improve the accessibility of such collections, increase the number of potential users, and increase the value of the information collection.

Research within Personal Information Management (PIM) has attempted to describe how people re-find information within their personal collections. These studies have typically focused for the most part on text collections, and usually conduct experiments using participants' own information collections. With a lack of a specialist user group, this study employs a methodology similar to that of the Data Mountain experiments [113, 114, 115]. Participants begin by organising a collection of unknown spoken word documents before completing re-find tasks. This has the advantage of revealing valuable information on organisation styles and strategies, as well as details on how spoken word documents are found.

Chapter 3 recommended that a manual spatial organisation tool that supported memory functions and provided better access to audio content would improve performance in re-find tasks. With this in mind, two versions of PAWS are compared (known as *basic* and *audio* for the remainder of the chapter). The main difference between the versions is the number of audio cues that are available. The primary aims of this study are to compare the times, error rates, and satisfaction levels between the two versions of PAWS. The role of organisation styles and strategies in relation to re-finding success are explored as well.

Features	***Basic audio***	***Extended audio***
Views	Workspace & map	Workspace, *AudioView* & map
Object type	Sound & landmark	Sound & landmark
Modality of views	Visual	Visual & audio
Audio cues	Simple playback	Simple playback, *hover-to-hear*, *multi-hover-to-hear*, *aspect-listen*, Landmark sound

Table 4.1: Overview the differences of features designed and implemented in PAWS.

4.2 PAWS Interface Features

Two versions of PAWS were used for this experiment. The design features of PAWS are those of the author. The difference between the versions is the number of audio cues available to users. Much of the implementation was carried out by experienced Java programmers. PAWS is written in Java 1.5 using JOAL [142], a Java wrapper for OpenAL 1.0 [143], for the audio implementation.

4.2.1 Visualisation of PAWS

The visualisation of PAWS is based on a combination of ideas from user- and computer-centred spatial tools (Sect. 3.1.1). Two views of a collection are available at all times: a workspace and a map (numbers 1 and 2, respectively, in Fig. 4.1). The workspace displays part of the entire available space (known as the universe) at full size. The map displays the universe in miniature. At all times, the map shows users where the working space is in relation to the universe. It is marked on the map by a green rectangle. The workspace view may be changed either by clicking and dragging on the workspace or within the green rectangle on the map.

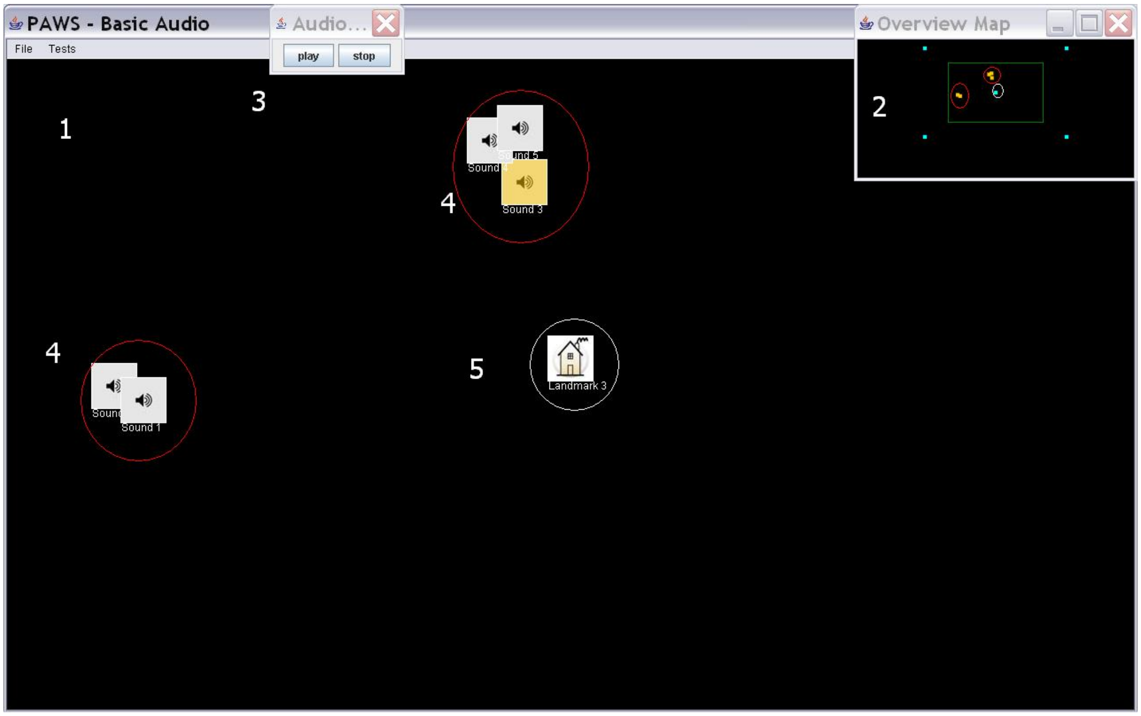


Figure 4.1: PAWS interface with limited audio features. Workspace, map, and basic audio playback control as well as one landmark are shown.

4.2.2 Object Types

Two types of objects exist in PAWS: a sound and a landmark object. Sound objects contain a spoken word document and are represented by a speaker icon (see Fig. 4.2) in the workspace and by an orange square in the map. Sound objects can be moved either by dragging them on their workspace or on the map. When a sound object is selected or hovered over, the icon changes colour. Sound objects can be positioned on top of each other in order to create piles. When sound objects that are positioned lower in a pile (lower Z-position) are clicked on with the left mouse button, it moves to the top of the pile (highest Z-position). Clicking with the right mouse button maintains the Z-position of sound objects. Number 4 in Figure 4.1 shows sound objects arranged in piles. The sound object with the orange background is the currently selected one.

Landmarks are a technique used within wayfinding to aid users in navigating around a space. This feature is based on other landmark techniques. For this experiment, five

Figure 4.2: Icon for a sound object.

Figure 4.3: Icon for a landmark object.

fixed positioned landmarks are represented by a house icon (see Fig. 4.3) in the working space and by a green square in the overview (see number 5 in Fig. 4.1).

4.2.3 Audio Features

The extended audio version has all of the visual features described above, and includes additional audio features. The two versions have been developed with regard to ensuring that there are no performance differences exist between the two. The initial interface for the extended audio version includes additional controls. An overview of the extended audio version is presented in Figure 4.4. The landmark button (number 3 in Fig.) opens the control panel for landmark sounds (see Fig. 4.5).

Simple playback

To listen to a sound object, the user clicks on it to select it upon which its background colour changes to orange- and then presses "Play" on the control panel. Selection works either by clicking on the sound object in the workspace or in map view. Only one sound is played at a time. If a sound object is selected and played while another one is playing, the first sound object is stopped, and the new selection is played. All sound objects play from the beginning. The user stops sounds by pressing "Stop" on the control panel (see number 3 on Fig. 4.1). All sound objects are numbered for the experiment (e.g., sound1, sound2), and changing their name is not possible . While in the extended audio

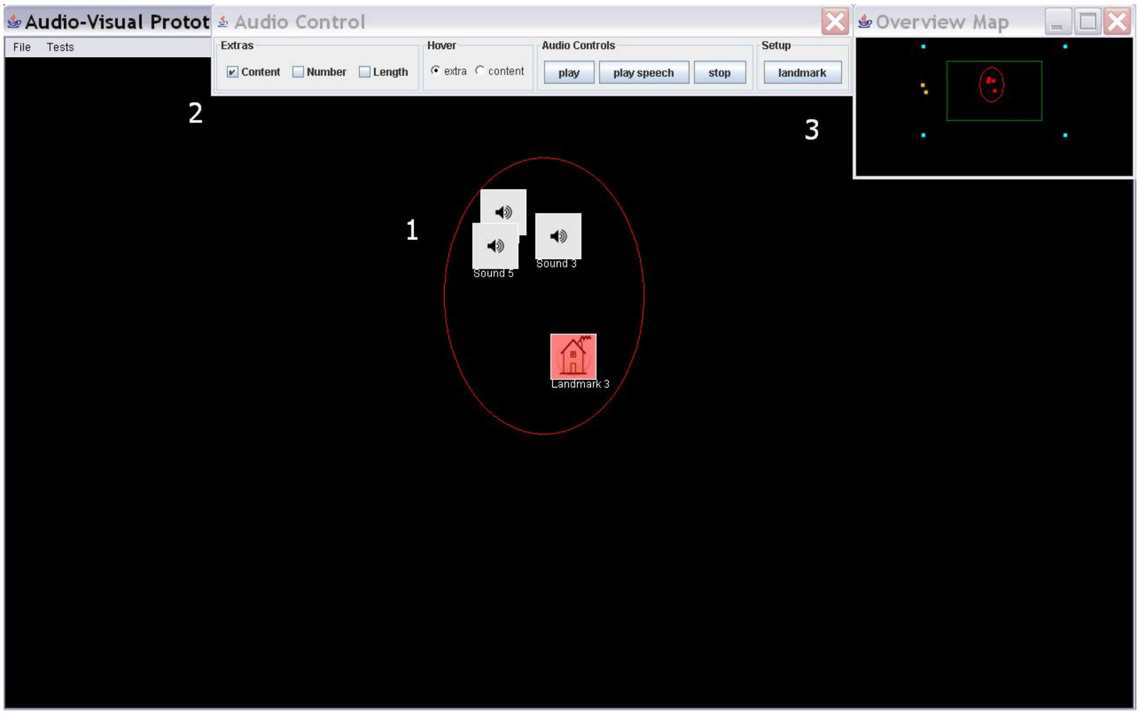


Figure 4.4: PAWS interface with extended audio features. Workspace, map and extended audio playback control as well as one landmark and three sound objects are shown. Red highlights that the landmark sound and the sound objects (marked in red in the map) which are currently being played simultaneously.

Figure 4.5: Landmark control panel for the extended audio playback features of PAWS.

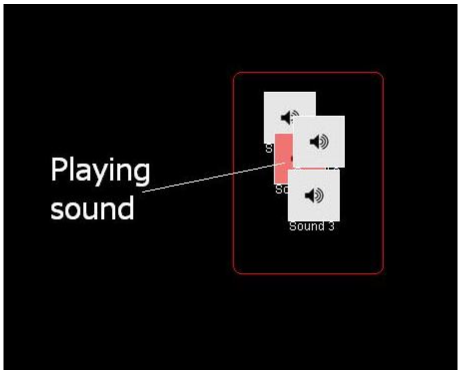
version, the position and loudness of the sound object depends on where it is located in the universe. The basic audio version does not show changes in loudness or direction. Simple playback the only audio feature available within the basic version.

AudioView

The *audioView* maps sounds from a 2D universe into a partial spherical 3D space (270 degrees surrounding a listener, not 360 degrees). The *audioView* aids users in distinguishing where a sound object is positioned within the universe. As it is possible to listen to sound objects without moving to their location (via listening using the map), it also provides an indication as to where users are in relation to the sound object that is playing. The loudest point in the universe is at the centre of current workspace. Section 3.2 highlighted tools that used spatial presentation of audio content for good effect. It is also recognised that it is easier to distinguish different sounds that are played simultaneously when the perceived distance and direction is different (see Sect. 3.2). To use JOAL's standard equations for positioning sound, 2D universe coordinates must be transferred into 3D coordinates. Standard mathematical equations were used to transfer the 2D coordinates into partially spherical 3D coordinates.

Hover-to-hear

When a sound object is hovered over with the mouse, after a delay of one second, the contents of the sound object start to play. This can be used in either the workspace or map view. The selected sound object changes its colour to red (both in the workspace and on the map) when playing. When the mouse moves away from the sound object, the playback stops. When sound objects are in piles, the top object in the pile, where the mouse may hover, plays when hovered over. Figure 4.6 shows sound objects are hovered over when in piles.



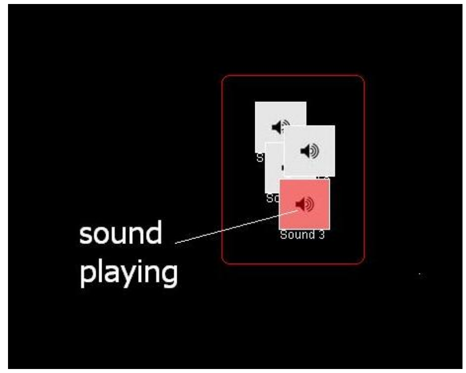


Figure 4.6: Hovering over sound objects causes them to change colour (from grey to red) and to play. When in piles, the sound object directly below the mouse cursor is played.

Aspect-hear

With the *aspect-hear* feature different aspects of a sound object may be listened to. An aspect represents one view, meta-data, or part of the content. In this case the different aspects are: content, number of voices, and length of sound object (see number 2 in Fig. 4.4). Prior to the experiment, the spoken word documents were tagged with meta-data that described the different values of each aspect. Using the control panel, users choose to listen to a certain combination of the content, number of voices, and length meta-data. As with visual attributes, audio attributes may be used to describe different aspects within information (see Sect. 3.2). For PAWS the length of an sound object is described as either short, medium, or long. The sounds used to represent this vary only in length (i.e., a short sound object is represented by a short sound). The number of voices is described as either one or many. The sounds used to represent this aspect are a single voice to represent a one-voice sound object, and many voices played simultaneously to represent a multi-voice sound object. The control panel for the extended audio version allows users to choose what is listened to at any time. When a sound object is hovered over, after one second, the contents (or chosen aspects) are played. The delay is intended to allow for the sound objects to be moved without listening to the contents. As with the simple audio version, users may listen to sound objects by selecting (clicking on) and

Figure 4.7: Aspect-hear audio control on PAWS.

pressing "Play" in the control panel. Depending on where the sound object is in relation to the loudest point, the loudness and direction change.

Audio Landmarks

Landmarks can have sounds attached to them. The idea is to make landmarks more memorable. Sounds may be attached to landmarks by the use of the landmark control panel (see Fig. 4.5). For this experiment, participants were able to select one out of a selection of nine sounds. It was also possible to not select any sounds for a landmark.

Multi-hover-to-hear

Multi-hover-to-hear combines *audioView*, *soundmarks*, and *aspect-hear*. When a landmark is hovered over, the landmark sound (if a sound has been chosen) and the five closest sound objects (chosen aspects) play. Presenting sounds simultaneously is hoped to reduce the time it takes to find an individual spoken word document (see Sect. 3.2). To improve the intelligibility of sounds, the *audioView* projects the sounds into a spherical space around the user (see description in *audioView*). As the centre of the workspace is the loudest point in the *audioView*, when objects are moved, the volume and direction of the sound and landmark objects change. The view is moved by clicking and dragging the landmark. This maintains the organisation of the sound objects in relation to the landmark, and allows users to identify which object is in the centre of the screen and, therefore, the loudest object. The map shows which sound objects are playing at any given time (see number 3 Fig. 4.4).

4.3 Methodology

The experiment was a repeated-measures design for independent variables: “version type”, with two levels (basic audio and extended audio) and “search type”, with three levels (known-item search, exhaustive search (audio property), and exhaustive search (content property)). Due to the risk of a learning effect between the two versions, the order the version type was presented in was counter-balanced; that is, all participants used both interfaces, but not in the same order. Half of the group used the basic audio version first, the other half of the group used the extended audio version first. Two sets of spoken word collections were used. Again, the order presented to the participants was counter-balanced. All participants completed the three search tasks in the same order.

4.3.1 Participants

Eighteen participants volunteered for the experiment. All participants were German speakers, whose English was sufficient for the experiment. Sixteen participants worked within the area of knowledge management, the other two were computer science students. All participants had experience with re-finding tasks from their daily life. None of the participants had experience with spoken word collections, but 17 had experience with organisation and search tasks relating to music (mainly MP3s). Only three participants were familiar with spatial organisation tools, although a further six reported to organise their computer desktops spatially. All reported to have normal hearing, although no tests were conducted to prove this. Participants were assigned randomly to one of the two groups.

4.3.2 Apparatus and Materials

Questionnaires

The **V**isual, **A**udio, **R**ead/write, **K**inesthetic (VARK) questionnaire tries to establish a person's learning preference [144]. Once known, learning materials can be designed that suit a particular learner and make the learning experience more effective. For example, visual learners learn more effectively when presented with visual materials, whereas audio learners prefer to use audio material, such as audio recordings and discussing ideas with others. Many people fall into a multi-modal category with particular leanings towards one or more of the other categories. That means the most effective materials cover all learning styles. The VARK questionnaire is not a standard learning preference questionnaire. However, it was suspected that those participants who had a particular leaning towards audio materials would perform better using the audio extended version. Therefore, the VARK questionnaire was used to give an indication if the participants were particular learning types. The test is short (only 13 questions) and easy and quick to administer.

Within this experiment, Likert questions were used to obtain opinions on the design of the two user interfaces [145]. In this experiment five-point Likert questions were used (e.g.,, 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree). The questions were asked at different stages during the experiment, and additional questions were asked for the extended audio interface. Table 4.2 shows the standard questions used for the two interfaces and Table 4.3 shows the additional questions for the extended audio version.

Audio Data

Two collections of 20 sound objects were created for the experiment. These sound objects came from BBC radio 2 interviews and required no previous knowledge in order for the content to be understood. The speakers discussed topics such as art, music, acting,

Task	Question
L1	"It was easy to place the sound objects."
L2	"I will be able to re-find sound objects quickly."
L3	"I remembered the location of the sound object needed."
L4	"The display has too much information on it."
L5	"Overall the interface is effective."

Table 4.2: Likert questions asked during the organisation and re-find experiment.

	Questions
AL1	"The additional sounds helped me to remember the location of the sound objects."
AL2	"The additional sounds are confusing."

Table 4.3: Likert questions asked for the extended audio features in the organisation and re-find experiment.

and architecture. Each sound object ranged from between 5 to 30 seconds, and had one or more speakers. The sound object collections contained a combination of male a female voices. All speakers used standard and clear English. Sound objects within each collection were presented to participants in random order. Both interfaces used both collections.

4.3.3 Procedure

Prior to the experiment, participants completed a VARK questionnaire and described their experiences with both spatial organisation tools and spoken word collections. If they had had experience with either, they described what tools they had used.

Before testing began, participants had time to familiarise themselves with the appropriate version of PAWS. First, by reading a set of paper instructions describing the relevant interface (see App. A.1 and App. A.2). To ensure participants all worked under

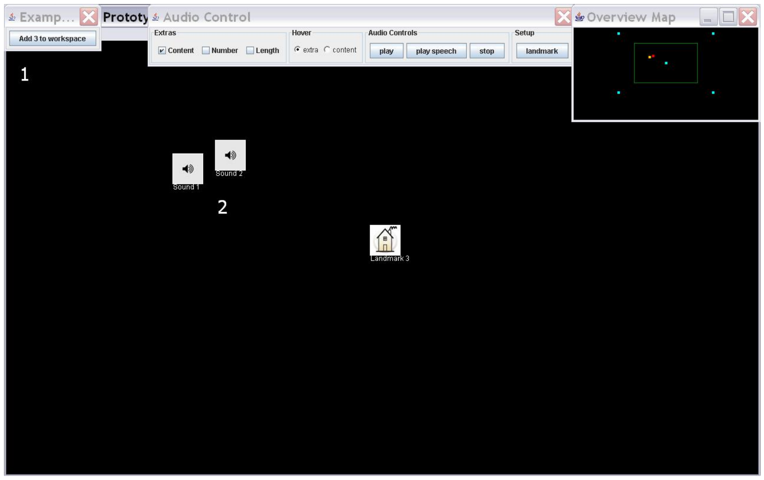


Figure 4.8: PAWS with extended audio features. Control panel for organisation task. Number 1 shows how to add sound objects to the workspace. Number 2 shows the added sound objects.

the same conditions, paper instructions for the relevant PAWS version were available at all times. Participants then tried out all the features of the version and completed a short training segment, which was a small version of the experiment. Participants organised five sound objects in their workspace. Each sound object was added to the workspace individually, where participants could move the sound object to a suitable position. Once satisfied with the placement, participants listened to the next sound object, and so on. Participants could listen to any of the sound objects as often as they wanted. Sound objects could be moved either using the workspace or the map. Figure 4.8 shows the training organisation task for the audio extended version. Number 1 shows the control panel that adds sound objects to the workspace when clicked on. Number 2 shows already added sound objects.

The known-item search, involved finding a specific single spoken word document. The wanted spoken word document was presented in sound. This task was completed once for the training. Figure 4.9 shows the control panel for the specific find task. Number 1 plays the sound that is wanted. Number 2 is clicked in order to select the sound the participant has found. Number 3 is clicked to complete the task and register the result.

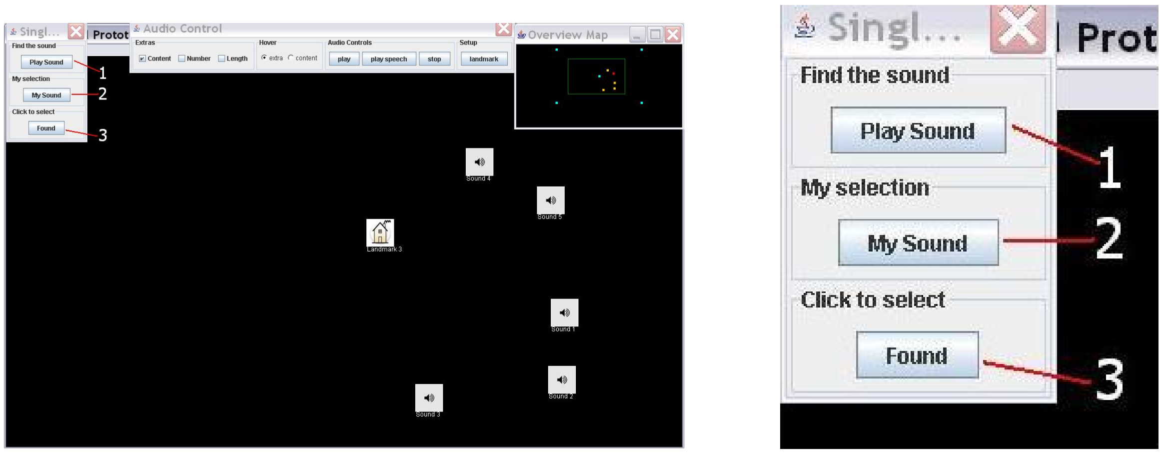


Figure 4.9: PAWS with extended audio features. Control panel for finding specific sound objects. Number 1 shows how to listen to the desired sound object, number 2 is pressed when the chosen sound is found. Number 3 is pressed to complete the task.

Both *exhaustive search* tasks involved finding a group of spoken word documents that were related to a text-based scenario. This task was completed once for training. Figure 4.10 shows the PAWS with audio extended features and the control panel used for both of the directed search tasks.

Participants were able to ask any questions concerning PAWS or tasks during training. When comfortable with both PAWS and the tasks, participants proceeded to the experiment.

Starting with a blank workspace, participants added spoken word documents to their workspace one at a time (20 in total). Participants could re-listen and move already positioned sound objects, using either the workspace or the map. The sound objects in the collection were presented to participants in a random order so as to reduce any effects of presenting information in a particular order. Before organising the collections, participants were asked to structure the spoken word collection in such a way that they would be able to re-find spoken word documents easily if required. The time taken for participants to organise the collection, as well as their final workspace were recorded for future analysis. After participants had organised their workspace, they were asked to

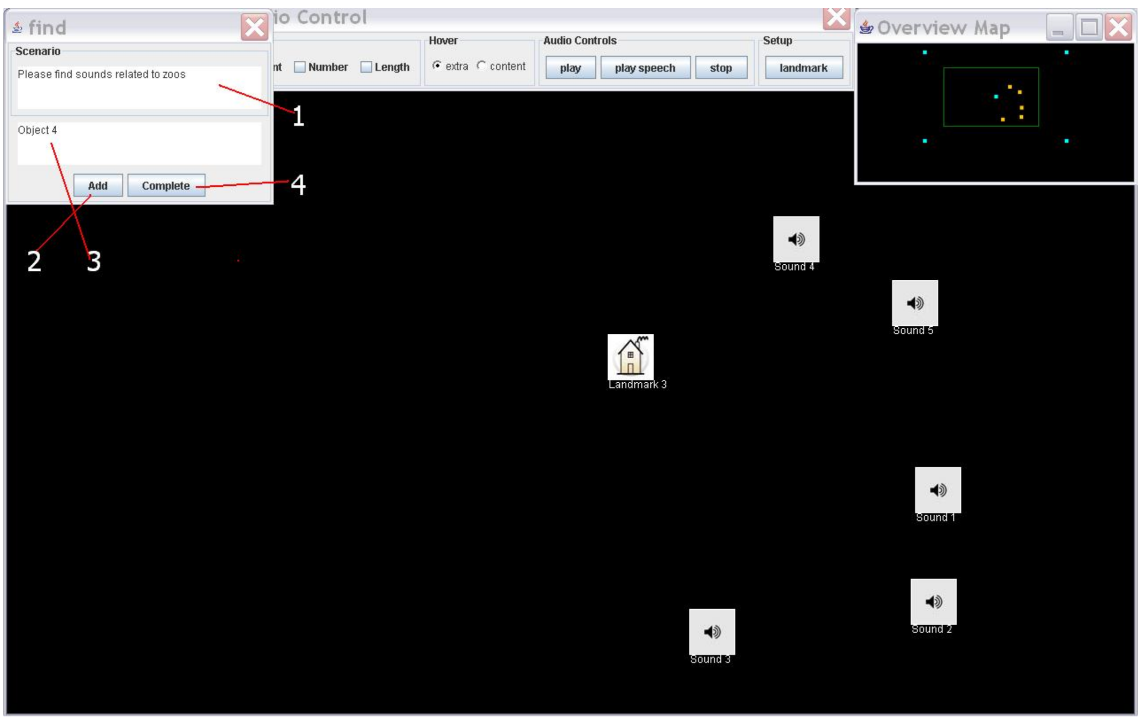


Figure 4.10: PAWS with extended audio features. Control panel for directed find tasks for both audio and content properties. Number 1 shows the written scenario, number 2 is pressed to select an appropriate sound. Number 3 shows where the chosen results are presented. Number 4 is clicked when all the relevant sound objects are found.

comment on how easy it was to position sounds, and how likely they were to re-find individual sound objects (Likert questions L1 and L2 from Tab. 4.2).

The *known-item search task* consisted of re-finding a given sound object as quickly as possible. The requested sound object was played to participants. This task was repeated five times, each time requesting a different sound object (the control panel for task 1 is shown in Fig. 4.9). The spoken word documents selected for retrieval were chosen at random in order to reduce the effects of one spoken word document being more memorable than another. Error rates and the times taken to find the correct spoken word document were recorded automatically.

Next, the *exhaustive search tasks* involved finding as many sound objects as possible related to a particular topic or audio property. For example, finding sound objects related to music or ones that feature female voices. The tasks were considered completed when participants felt that they had found all relevant sound objects. The exhaustive search task was repeated five times (twice for audio property, three times for topic property). The time taken and the number of sound objects found were recorded for future analysis (Fig. 4.10 shows the interface for tasks 2 and 3).

After the experiment, participants were presented with the remainder of the Likert questions (Tab. 4.2) and were also asked to describe what organisation strategy they had used. Those participants who were being tested with the extended audio version were asked the Likert questions described in Table 4.3. Additional comments concerning PAWS were given, as well as version preferences.

4.4 Results

The results for this experiment were obtained using a combination of automatically recorded information (i.e., time and error rates, saved workspaces), questionnaires (i.e., VARK and Likert), interviews, and observation. The main findings are presented below.

Basic audio	**Extended audio**
8m 39s (SD 3m 5s)	9m 1s (SD 4m 14s)

Table 4.4: Storage times and standard deviation for basic and audio extended interface.

Basic audio	**Extended audio**
8m 59s (SD 3m 30s)	10m 5s (SD 3m 24s)

Table 4.5: Storage times and standard deviations for first and second attempts at organising collections.

4.4.1 Organisation Task

Timed

The time each participant took to organise 20 spoken word documents was recorded. Table 4.4 shows the mean time, and standard deviation, in minutes to organise the basic and extended audio interfaces. A paired samples t-Test [146] was conducted on the version type and the time taken to organise. Although participants were quicker when organising using the basic audio version, there was no significant main effect ($t(17) = -0.29$, $p = 0.78$).

As participants had completed this task on both interfaces, the time taken to organise for both first and second attempts was noted. This is shown in Table 4.5. A paired samples t-Test was conducted on the time taken to organise and the order they had organised in. Participants took significantly longer to organise a set of 20 sound objects in their second attempt ($t(17) = -2.59$, $p = 0.02$). Again, participants were slower to organise while using the extended audio version, but not significantly so.

Strategies

Participants had a number of styles and strategies for organising their workspaces. These can be divided into three decisions: use of space, property to organise by and alignment

Dimension	Possible values
Use of space	viewport — whole space
Alignment	tidy — messy
Property	topic, audio, both, none

Table 4.6: Axis of possible variations of organisation styles.

(see Tab. 4.6). Observations also indicated that some participants preferred to add the collection to the workspace prior to organising, while others organised it while adding the sound objects. This meant occasionally updating the organisation to suit the recent additions. The strategy used to organise a collection can be determined by mathematical calculation. However, in this case, observation and interview methods were used to determine participant preferences.

Use of space went from using only the viewport space, so all spoken word documents can be seen at all times, to using the whole space. Often, participants grouped sound objects around landmarks, so categories would form. Using only the viewport meant that all objects were visible at all times, but gave less space to each category. Using all of the space meant time may be spent moving around the space, but categories got more space (see Fig. 4.11). Another organisation decision was which property to organise by. Observations and interviews showed that users grouped by either topic, audio properties, or a combination of the two(see Fig 4.12). Alignment refers to how straight the lines are between spoken word documents. Some preferred straight lines at equal distances, others preferred a more “messy” approach (see Fig. 4.13).

In an attempt to solve the difficulties of positioning sound objects that belonged in more than one category they were positioned in between all relevant places. For sound objects that did not seem to belong to one obvious place, a miscellaneous pile was created. If another sound object appeared that was related to a sound object in a miscellaneous pile, a new category was created. Many participants preferred to re-

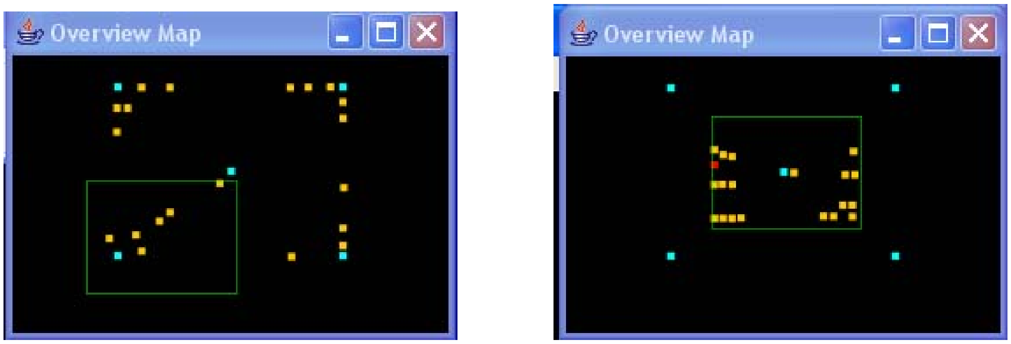


Figure 4.11: Use of space dimension. An example of a whole space (right) and a work space (left) organisation.

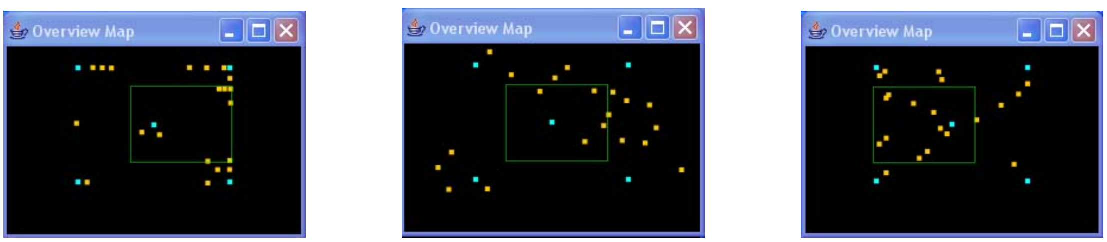


Figure 4.12: Property dimension. Examples of topic (right), audio (centre), and both (left) organisation strategies.

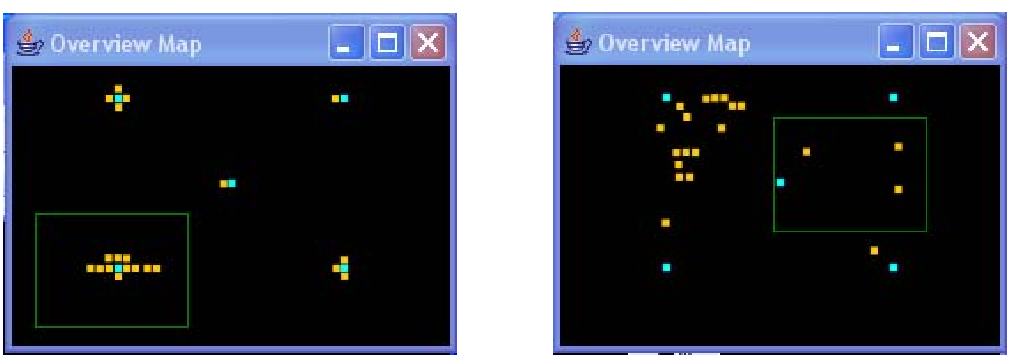


Figure 4.13: Alignment dimension. An example of a tidy (left) and “messy” (right) organisation.

	Basic audio	**Extended audio**
Time	27s (SD 11s)	36s (SD 20s)
Error	1	1

Table 4.7: Times and error rates for task 1 for basic and audio enhanced interface.

organise their workspaces when necessary. For many participants, space was used to convey extra details about a sound object. Half of the group partially changed their strategy for the second attempt of the experiment. For example, they changed from using the universe to just using the workspace. This decision by so many of the group to change one or more axes in their organisation style, implies that their original attempt was not satisfactory.

4.4.2 Known-item Re-find Task

In the known-item re-find task, participants were asked to re-find a specified sound object. They were given an audio cue of the sound they needed to find. Each participant repeated this task five times. The result for the time and error rate was an average of their five attempts. These times and error rates are given in Table 4.7. The Mean times are given in minutes, and the mean error rates are expressed in percentage of correct answers.

A repeated measures t-Test was performed on the version type and task time. Participants were significantly faster at finding specific sound objects with the basic audio version ($t(17) = -2.37$, $p = 0.03$). The error rate was the same for both interfaces. A t-Test was also conducted to compare the time taken to complete the tasks between first and second attempt. No significant difference was found. However, participants were faster in their second attempt (29 seconds (SD 10 seconds)) than their first attempt (36 seconds (SD 21 seconds)).

	Basic audio	**Extended audio**
Time	1m 26s (SD 1m 11s)	1m 36s (SD 41s)
Number found	4.04 (SD 0.81)	3.1 (SD 0.8)
Average time	22s (SD 20s)	35s (SD 21s)

Table 4.8: Times and number of spoken word documents found, and average time to find an spoken word document for task 2 for basic and audio enhanced interface.

4.4.3 Exhaustive Search by Topic Property Task

Participants were asked to find spoken word documents related to a given topic, such as music. An answer was considered correct if a keyword appeared in the spoken word document that related it to the given topic. Each participant completed the task three times, and was given an average result for both the task time and number of correct spoken word documents found. The average time to find a correct result was computed from the task time and the number of correct spoken word documents found (see Tab. 4.8).

A repeated measures t-Test showed no significant difference in the time between the two interfaces ($t(17)= -0.58$, $p = 0.57$). However, significantly more correct spoken word documents found in the basic interface than in the audio enhanced interface ($t(17)= 4.28$, $p = 0.001$). A repeated measures t-Test was also performed on the average time to find an object. Although completing the task with the basic audio version was quicker (22 s) than with the audio enhanced interface (35 s), this was not a significant result ($t(17) = -2.05$, $p = 0.06$). Also, the standard deviation for completing the task was lower for the extended audio version (41 s) than the basic audio version (1 m 11 s). This suggests that even though there were no significant differences in the time taken to complete the task, there was more variability within the basic audio group.

	Basic audio	Extended audio
Time	1m 47s (SD 1m 27s)	1m 40s (SD 48s)
Number found	4.8 (SD 1.8)	4.25 (SD 1.5)
Average time	27s (SD 16s)	27s (SD 20s)

Table 4.9: Times, number of spoken word documents found and average time to find a spoken word document for task 3.

4.4.4 Directed Search by Audio Property

Participants were given a textual description instructing them to find as many spoken word documents as possible related to a given audio property, such as a voice type. Participants completed the task twice with times and error rates recorded for both attempts. The results for both the time and error rate were averaged. The average time to find a correct spoken word document was calculated by dividing the number of spoken word documents found by the average time (see Tab. 4.9 for details).

A repeated measures t-Test was performed on the version type and time, the version type and the number of sound objects found, and the version type and the average time. No significant differences were found between the two versions. The standard deviation for the time taken to find relevant sound objects is smaller for the extended audio version, indicating less variability in performance between participants.

4.4.5 Use of workspace vs Re-finding Tasks Performance

Use of space

A Pearson correlation was performed on the use of space and time to complete known-item task and the average time to find a spoken word document in both of the exhaustive search tasks. No significant results were found relating how the space was used and average time to find a spoken word document. However, the less space used, the less time it took to find a spoken word document. It also suggests that 20 sound objects may not have been enough for a space larger than a monitor to be useful.

Organise by topic/audio/both

Calculations were conducted on how the organisation style and known-item search time, and both of the exhaustive search average times correlated. A correlation was found between organisation style and the exhaustive search (audio property) average time (Kendalls Tau -.27, p=.05), in favour of those participants who organised by taking into account both the topic and audio properties. This indicates that organising using as many dimensions as possible speeds up any re-finding later on.

Organise by Alignment

No correlations were found between alignment and search time in any of the search tasks. This dimension appears to match user preferences more closely.

4.4.6 Subjective Measures

The discussed subjective measures were gathered using Likert questions, observation, interviews, and general comments from participants. Non-parametric statistics were calculated of the Likert questions. No statistics were calculated for information gathered from the observation, interviews, and general comments. These types of information were intended to better understand what had happened in the experiment, and to improve PAWS.

Likert Questions

During each interface, participants answered five Likert questions in order to gauge the satisfaction levels with the interfaces. A 5-point Likert scale was used (from 1 - strongly

Question	Basic audio	Extended audio
L1 "It was easy to place the sound objects"	3.28 (SD 1.13)	3.39 (SD 1.20)
L2 "I will be able to re-find the sound objects quickly"	3.28 (SD 0.89)	3.44 (SD 0.86)
L3 "I remembered the location of the sound object needed"	3.22 (SD 0.88)	3.78 (SD 0.81)
L4 "The display has too much information on it"	1.33 (SD 0.49)	1.83 (SD 0.86)
L5 "Overall the interface is effective"	3.67 (SD 0.97)	4.11 (SD 0.58)

Table 4.10: Summary of the results for Likert questions for Experiment 1.

disagree to 5 - strongly agree). The results show the average Likert figure for each interface type. The mean of the results are given in Table 4.10.

Wilcoxon tests were performed on each question [147]. LI and L2 showed no significant results. However, the results from L3 and L4 are significant for the extended audio version ($T = -2.153$, $p = 0.03$) and ($T = -2.041$, $p= 0.04$). The result for L5 is also close to being significant ($T = -1.734$, $p= 0.08$). A significant number of participants (17 participants) reported they preferred using the interface that had additional audio features.

Feedback for PAWS

The feedback on using PAWS was encouraging. All participants reported it was easy to use and a useful way to organise and re-find spoken word documents. Most comments fell into the categories of landmarks, *multi-hear*, annotation, and organisation.

Landmarks. Landmarks were used often as a place where groups of sound objects were assembled (clusters). In the extended audio version, a landmark could have different sounds attached to it. This was generally not liked nor found useful by the participants. They felt that the attached sounds had no additional meaning to the landmark, and gave no extra benefit. One suggestion for improving the aural memorability of landmarks, was that the sound the landmarks attached had to be *borrowed* from the sound objects that were closest to the landmark. The assumption being that sound objects that were more representative of a cluster would be positioned closer to a landmark than those that were less representative. If the landmark had the same sound as that of its closest (and arguably) most related sound object, then it would better describe its cluster. It was also suggested that this way of collecting an identity was only available on the map.

Multi-hear. In general the number of simultaneously played sounds was too high for most users. In order to make use of this function, the space between objects had to be larger, and this could detract from the organisation. An improvement for this feature would be to reduce the number of sounds playing simultaneously, and to make the differences in volume and location of the sounds more distinct, thus making the individual sounds appear more separate and distinguishable. In general, approximately one fourth of participants used this feature in order to browse for information in the exhaustive search tasks.

Hover-to-hear. The participants appreciated the hover feature, but felt the delay was too long. Other recommended improvements were to allow fast-forwarding and rewinding of the sound objects.

Organisation. Many participants felt that two dimensions was not adequate. Suggestions for improvements included additional visual cues, such as colour, size, and shape. Others felt that being able to annotate both landmarks and sound objects would have increased their success in re-finding information. In the extended audio version, some participants reported to having spent less time to organise their collections, as they felt that the browsing support offered would enable them better to re-find wanted information more easily. Another suggestion was to be able to change the view of information to suit the current user's needs. Some participants reported remembering when they had heard a sound object, especially sound objects near the beginning or at the end. This implies that supporting a temporal structure could offer benefits.

4.5 Discussion

4.5.1 Organisation Styles and Strateiges

As speculated, different organisation strategies were apparent. Three dimensions were identified that were important while organising the collection: use of space, organised property, and alignment. Piling and clustering were also common techniques for organising related sound objects (also found in Section 3.1.1). More than half of the group changed one or more of these dimensions during their second attempt of organisation. This may indicate that they felt that they had not found an optimal way of organising the collection. It is suspected that the first attempt was a clearer indication of how they organised information naturally, whereas the second was more geared towards spoken word information. It can be assumed that with more practice, different organisation techniques would emerge. There is an indication that participants felt that organisation was important to successfully re-find a sound object, as there was a significant increase in time spent between the first and second attempts (see App. A.4). However, the cost, in terms of time spent, in organising such a small collection was relatively high. This may mean that using a user-centred approach to organisation may only be beneficial for those people who use organisation as a way of analysing data (e.g., sense-making tasks). Ways to reduce the time to organise while maintaining the benefits of user-centred organisation include supporting users in organising a small portion of their collection, so they then organise the remainder based on this organisation, to users describing basic rules and roughly organising the collection. An improvement in navigation cues could reduce the need for an effectively created organisation.

In order to test further, a comparison of manually organised, partially organised and automatically organised collections could better identify the role of organisation within re-finding in personal collections. The choice of organisation did not appear to be dependent on the learning style.

4.5.2 Re-finding Tasks

It was hypothesized that participants would be significantly faster and make fewer errors while using PAWS with extended audio. However, the participants performed significantly better in re-finding a specific item in the basic audio version. Reasons for the increase in time spent using the PAWS extended audio version could be due to a lack of experience with such tools, a lack of experience with spoken word documents, and the fact that audio is a slower medium to process. The increase in time may be due to the audio features not being optimal. Another reason for the slowness in extended audio is that the participants used well-established finding techniques, and when those failed, tried to use the extended audio features. This would naturally make the extended audio version slower, as the features were being used as a last resort. The error rates were the same for both versions.

It was hypothesised that the performance in terms of both time and number of correct sound objects found would be better for the PAWS extended audio version. No significant difference was found for either factor between the two versions. A reason for this could be that similar search strategies were employed in the two versions. For this task, the selected organisation style made the difference.

It was hypothesised that the performances in terms of both time and number of correct sound objects found would be better in PAWS extended audio version. There was no significant difference found between the two versions for either performance measurement. However, the standard deviation was lower for the extended audio version. This suggests that although the average time was approximately the same for both versions, those participants with less successful finding techniques performed better with the extended audio version than with the basic audio version. This indicates that the provided audio cues added value to those participants who had difficulties in completing this task.

In each case, a further test comparing experienced users with non-experienced users could help to confirm whether the favour towards the basic interface in terms of performance was due to a learning difficulty, or if the audio features added unnecessary complexity to the interface. Again, learning preferences indicated by the VARK questionnaire did not appear to affect performance within the tasks. A study by Grohn et al. [148] who compared a visual-only, an audio-visual, and an audio-only version, showed that the audio-visual interface out performed the other two.

Common strategies used to re-find sound objects was to first look in the location where it was expected to be (location-based search). For those participants who had a less organised space, it meant that finding the right area could be problematic. When looking for a sound object, participants listened briefly (a few seconds) to each sound object. Through observation, participants who could connect attributes about the voice with the content had to listen to a potentially relevant sound object for less time than participants who could not remember any of the details. Naturally, the task of listening to an object in order to verify it meant that re-finding in audio was slower than if it had been a visual collection. A similar study using the Data Mountain showed that the average time for most participants to complete a known-item re-find task was 5 seconds (organised dataset of 33 webpages). For the similar task in this experiment, the average time was 27 seconds for the basic audio, and 36 seconds for the extended audio version. These comparative results between webpages and spoken audio help to demonstrate that audio information is much slower to re-find. Memory was an important factor in re-finding. The act of manually organising a collection seemed to aid participants in re-finding sound objects later. Participants' descriptions of how the found wanted items included, "I heard it recently, so it must have a high number in its name", and "I know it's in a large group". The correlation between the time taken to organise in the second attempt and an improvement in re-finding specific items suggests that organisation does improve memory cues.

4.5.3 Subjective Satisfaction

It was expected that participants would be more satisfied using the PAWS extended audio version. Although significantly more participants reported to prefer using the extended audio version, the Likert questions show little difference between the two versions. Seventeen of the participants reported they prefered using the PAWS with extended audio. Reasons for this preference included more ways in which to search for information. Hover-to-hear was a popular feature, although the delay was annoying and, in some cases, participants found using the control panel to play back was quicker. The Likert questions did not show a significant preference for the extended audio version, although individual questions highlighted that participants were more confident in finding information using the audio extended version. The preference may be due to the newness of the interface and the extra dimensions offered. Although, the basic audio version was commented as having only rudimentary visual cues, it made the increased possibilities of the extended audio more appealing.

4.6 Chapter Summary

This chapter presented an organisation and re-find experiment using PAWS. PAWS was designed by first capturing user needs while re-finding personal information (see Chap. 2), and then analysing tools that supported these behaviours (see Chap. 3). Therefore, the tool is designed to support organisation and re-finding within a personal spoken word collection.

The main findings of this experiment were that memory cues, available while organising information, indicated to have played an important role in re-finding information. Although the basic audio interface performed significantly better in some situations, participants, on the whole, preferred to use the extended audio version. Promising features of the extended audio version were the ability to listen to more than one spoken word document simultaneously (*aspect-hear* with no landmark sound attached, and only content selected) and *hover-hear*. The additional audio features of aspects and audible landmarks were not used successfully, nor appreciated. The increase in time taken in the extended audio version may be due to the fact that audio is a slower medium, and that incorporating audio where it is not strictly necessary increases the time it takes to re-find information. The next chapter presents an experiment on seeking tasks using PAWS II. PAWS II was developed based on the findings of this chapter.

Chapter 5

Seeking Experiment

Chapter 4 described PAWS and an experiment that focused on organising and re-finding spoken word documents using two versions of PAWS. The main findings showed that, although organisation acted as a useful reminder for re-finding spoken word documents, the time taken to organise a small collection was high. In such cases, it is expected that users would prefer to not organise a collection in this way, unless the organisation provided some added value (e.g., sense-making tasks).Evidence gathered shows that if the perceived cost of maintaining a collection (e.g., organisation, changing meta-data) was high, the collection would not be maintained. Therefore, PAWS II is aimed at the *automatic organisation* user group. Promising features in PAWS were the ability to listen quickly to a spoken word document (*hover-to-hear*) and listening to spoken word documents simultaneously (*multi-hear*). Different playback and navigation features have been designed for PAWS II based on the findings from Chapter 4. PAWS II is designed to be used as an interface for both personal and public collections. In this experiment, tasks typically associated with public collections are tested using three versions of PAWS II (named *basic, ear,* and *navigation*). The main difference between the versions is the playback and navigation features.

The remainder of this chapter is organised as follows: Section 5.1 describes the motivation for the experiment, relating it to previous studies. Section 5.2 describes the features of the three versions of PAWS II. Section 5.3 describes the methodology that was used. Section 5.4 gives the results of the experiment. Section 5.5 describes the main findings of this research and Section 5.6 concludes the chapter.

5.1 Motivation & Background

This experiment aims to replicate seeking behaviour for known-item, exhaustive and exploratory seeking tasks within public spoken word collections for the *automatic organisation* user group. Due to the difficulties of finding a group of users with a personal spoken word collection, the collection used in this experiment is unknown to all participants. It is expected that PAWS II will work in a similar manner for PIM seeking tasks. Time, error rates and paths chosen in the information, as well as the use of features are recorded for analysis. Three versions of PAWS II were tested: a *basic* version with minimal audio and memory support, an *ear* version enables listening to more than one spoken word document at once and a *navigation* version where navigation techniques are supported. For the remainder of this chapter the three versions are called *basic, ear,* and *navigation*. The goal of this experiment is to discover if the additional features provide improved access to spoken word collections, and to more accurately describe seeking behaviours within a spatially organised spoken word collection by using the information paths created by participants. The primary aims of this study were to investigate if the additional audio features, and navigation features improved performance in terms of time and error rate. An investigation of seeking strategies, use of designed features, and listening strategies are included. It is expected that the additional features will improve access to individual spoken word documents, thus improving performance and satisfaction levels.

5.2 PAWS II Features

5.2.1 PAWS and PAWS II

The features of PAWS II were inspired by ideas explored in Chapter 3,based on expected seeking behaviours (see Chap. 2) and the feedback and results described Chapter 4. It is designed for both professional and casual users, and for all types of information (ephemeral, working, and archive). The most promising features of PAWS were the ability to listen to the audio content quickly (hover-to-hear), simultaneous presentation of sound, multiple views, and easy methods with which to explore the views. The cost of organising a large collection for some tasks can be considered to be too high when doing it manually. A difference in PAWS II is that the organisation is created automatically. Landmarks were frequently used as a clustering method in PAWS. In PAWS II manual organisation, and the ability to move individual sound objects is not possible. Therefore, in PAWS II landmarks, as a grouping device, is replaced with clusters that are created automatically. The main aim of PAWS II is to support personal collections. However, it would also be possible to use this tool for public collections. A comparison of the main features of PAWS and PAWS II is summarised in Table 5.1.

5.2.2 Implementation Details

The design features of PAWS II are those of the author. Much of the implementation was carried out by experienced Java programmers. PAWS II is divided into three layers: user interface; model; and storage mechanism. It is implemented as a monolithic system, but this structure could be divided into three layers. PAWS II is written in Java 1.5 using JOAL [142] for the audio implementation. JOAL is a Java wrapper for OpenAL 1.0 [143]. Java serialisation was used as a storage mechanism, but others may be used. The model is based on other user-centred spatial organisation tools.

Features	**PAWS**	**PAWS II**
Organisation	Manual	Automatic
Grouping	Landmarks	Clusters
Views	Workspace Map	Workspace Map
Exploring information	Panning	Panning, Zooming
Listening	*Hover-to-hear*, Control panel listen, Landmark-listen	*Hover-to-hear*, *Click-to-hear*, *AudioSphere*
Memory devices	Organisation	*Timeline*, *IListen*, *Prediction*

Table 5.1: Comparison of PAWS and PAWS II features.

5.2.3 PAWS II

Three versions of the interface were used in this experiment (named: *basic, ear,* and *navigation*). All interfaces shared the graphical features of the basic interface. The *basic* interfaces provides minimal audio and navigation possibilities. The *ear* interface includes all features from the *basic* interface and allows more than one spoken word document to be listened to at once. The *navigation* interface includes all features from the *basic* interface and allows participants to move back and forward within their history and to ask for a prediction. Table 5.2 provides an overview of the features of the three versions of PAWS. The features are described in more detail below.

Visualisation

Two views were available on the collection: a workspace and a map. The workspace displayed a portion of the possible collection (universe)(number 1 in Fig. 5.1). The map displayed the entire universe in miniature (number 2 in Fig. 5.1). The white rectangle in the map shows where the workspace is in relation to the universe.

Features	***Basic***	***Ear***	***Navigation***
Views	Workspace & map	Workspace & map	workspace & map
Navigation	Pan & zoom *iListen*	Pan & zoom *iListen*	Pan & zoom, *iListen* *timeline*, *prediction*
Listening options	Click-to-play/stop	Click-to-play/stop *AudioSphere*	Click-to-play/stop

Table 5.2: Overview of features in the three versions of PAWS II.

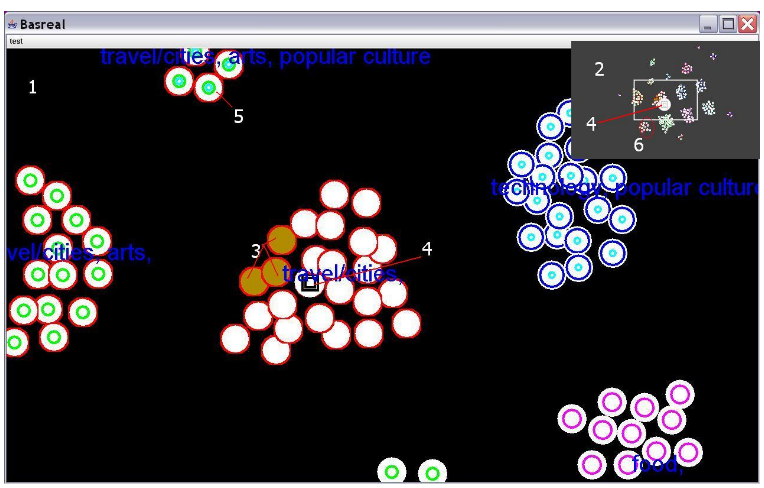


Figure 5.1: Basic interface showing the workspace, map and spoken word objects. Listen indicators that are white have not been listened to, listen indicators that are orange have been listened to.

Sound Objects

All spoken word documents are represented as a circle in the workspace (see Fig. 5.1). Each circle has two elements: a background colour and either one or many coloured circles. The background colour shows the state of *iListen*. If it is white, the spoken word document has not been listened to; if it is orange, the spoken word document has been listened to (compare numbers 3 and 5 in Fig. 5.1). *iListen* is visible in both the workspace and the map. No difference in colour is apparent to show how long a spoken word document has been listened to. The coloured circles represent what categories a spoken word document belongs to. One circle signifies that the spoken word document belongs in one category, two circles signify that it belongs in two categories, and so on. For this experiment, a spoken word document could belong to a maximum of three categories. When a spoken word document is clicked on *iListen* changes colour and its contents play (click-to-play). Clicking on an empty space within the workspace stops the playback. If a different spoken word document is clicked on while another is playing, playback of the previous document is stopped and the new spoken word document is played. Playback also begins at the start of the spoken word document's contents.

Automatic Organisation

Different techniques exist to automatically organise a collection based on pre-defined rules (see Chap. 3.1.1). A number of difficulties exist when organising collections. In the previous experiment, it was apparent that a number of organisation strategies were used, and the results showed that no significant correlation between organisation strategy and re-find performance existed; this would indicate that organisation strategies are essentially refined to suit the needs of an individual. Therefore, creating an organisation that suits all individuals is probably not possible. Instead, PAWS II provides a semi-organised collection based on category similarity, and additional audio and navigation cues to compensate for the lack or personal organisation.

The chosen spatial organisation algorithm chosen is based on the idea of positioning the objects that have the most categories in common closest to one another. As such, the algorithm is based on a concept of "category similarity", which can be reduced to a numeric value between 1 and 0. The value 1 represent two spoken word documents with an identical category set, and the value zero represents two spoken word documents that do not share a single category. The value of intermediate combinations can be calculated in many ways. The important aspect is that "category similarity" reflects the number of categories being shared. One way to do this is to count the number of shared categories and divide them by the number of categories that could have been shared. The algorithm is divided into three phases and is presented below.

Phase 1 **Initialisation.** Sound objects are positioned randomly within the available universe (see left image on Fig. 5.3).

Phase 2 **Iterative moving of sound objects to categories.** Each sound object is moved around the universe iteratively controlled by the following prioritised rules (see middle image on Fig. 5.3).

Rule 1 **Avoid any overlapping spoken word documents.** Move spoken word documents directly away from overlapping spoken word documents by half of a sound object's diameter. Test against sound objects in the whole universe.

Rule 2 **Avoid being near sound objects of different categories.** Move sound objects that have a different category signature directly away from other sound objects found within a circle area.

$$repulseFact\left(1 - \frac{distance}{circleDiameter}\right)(1 - catSim)\,\frac{averageSiblingCount}{siblingCount} \qquad (5.1)$$

repulseFact: Repulsion factor - constant adjusting the power of the rule

distance: straight line distance found between the centres of the two sound objects

distance: distance of 12 sound objects diameters

catSim: Category Similarity - calculated ratio between 1 and 0

AverageSiblingCount: average number of sound objects, within the universe that has identical categories

siblingCount: number of sound objects with identical categories to the present sound object

Rule 3 **Attract pods of similar categories.** Move pod directly towards any other pod in the universe that has the same category.

$$AttractFact \cdot distance \cdot catSim \cdot \frac{averageSiblingCount}{SiblingCount} \tag{5.2}$$

AttractFact:Attraction factor - constant adjusting the power of the rule

distance: straight line distance found between the centres of the two sound objects.

catSim: Category similarity - calculated ratio between 1 and 0

averageSiblingCount: average number of sound objects within the universe that have identical categories

siblingCount: the number of sound objects with identical categories to the present sound object

Phase 3 **Organise items of groups by iteration.** The final phase uses the same algorithm as phase 2. The reason of introducing a third phase is to reduce the strength of the attraction factor (rule 3) so the anti-collision rules 1 and 2 become dominant. This is achieved by dividing the attraction factor of rule 3 by a value of 1000. Without this change in attraction factor, we would either see individual items getting isolated

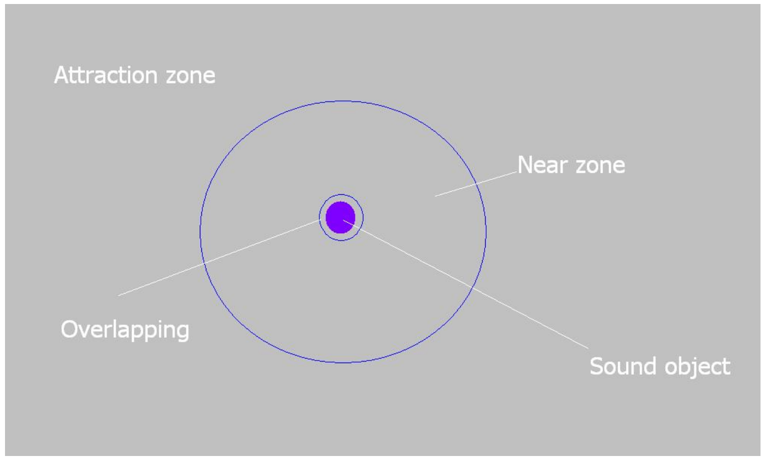


Figure 5.2: The gray area is where the attraction force is applied. All objects within the large blue circle lie within the near zone. The small blue circle shows the area that sound objects should be repulsed from.

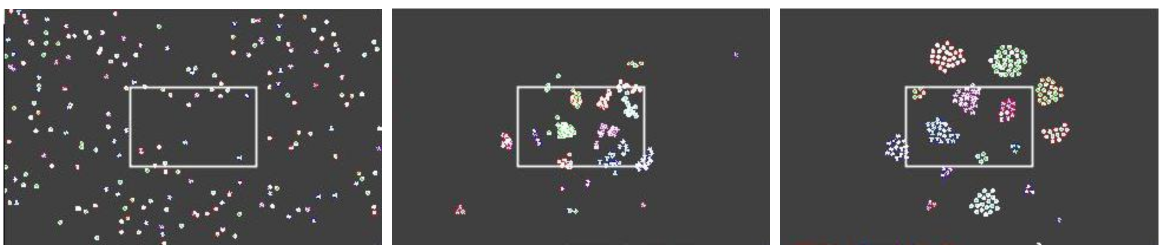


Figure 5.3: Stage one (left):adding all spoken word documents randomly to the workspace. Stage two (centre):clusters of related topics begin to form. Closely related clusters are positioned closer to those that are less related, where possible. Stage three (right): completed organisation. Objects move to reduce an overlapping of space.

from similar items, or alternatively, with a very high attraction factor, the shape of the formed groups would become distorted.

Navigation Support

Panning. Panning is frequently supported within spatial organisation tools 3.1.1. It allows the viewed space to be moved without requiring a change in size of the objects. To pan, click-and-drag on either the workspace or within the white rectangle in the map, and move to the desired position. To move to a different part of the universe without

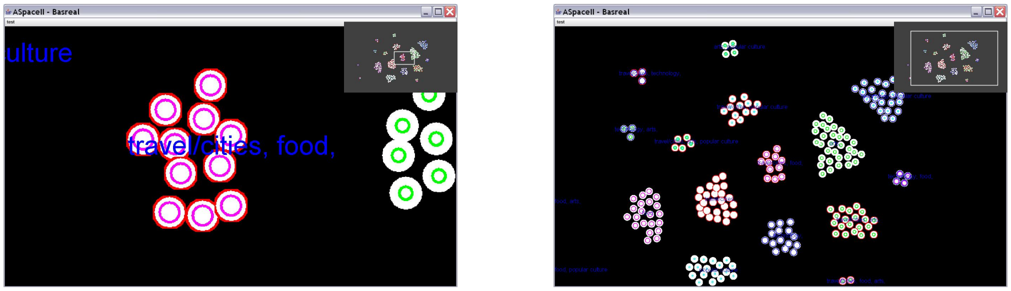


Figure 5.4: Basic interface showing a zoomed in and a zoomed out display.

dragging, click anywhere on the map. The clicked position becomes the centre of the workspace.

Zooming. Zooming is frequently implemented within spatial organisation tools 3.1.1. It allows users to increase or decrease the amount of visual information available. To zoom, use the scroll wheel on the mouse (or "+" and "-" on the keyboard). Figure 5.4 illustrates zooming within PAWS II.

Timeline. The *timeline* feature allows users to move backward and forward between sound objects that have already been listened to. This is intended to offer similar support to the use of the backward and forward buttons within browsers. When a user moves backward, a red line on the map shows the last five objects on their timeline. As it is not necessary to move to the spatial position in order to listen to a sound object, users may listen without moving. However, if users wish to investigate the cluster that the present sound object is in, they can press "enter" to move them to the position of the sound that is playing. Figure 5.5 shows the *timeline* feature of the navigation interface.

Prediction. A prediction is based on previous relevant choices. A relevant choice is determined by how often objects are listened to, how long they are listened to, and what clusters have been listened to. Difficulties arise when users listen to an object for a short period of time. This could be interpreted as either a definite no or a definite yes. To

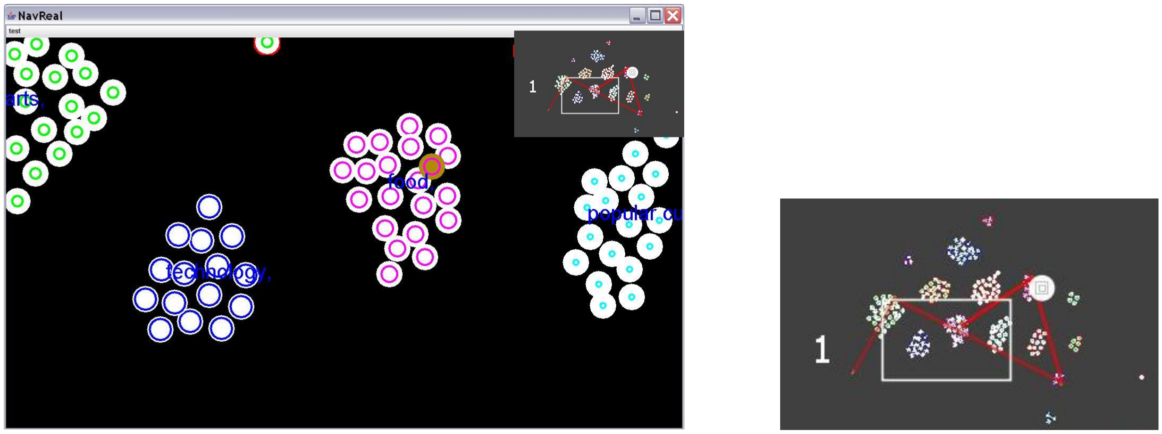


Figure 5.5: Timeline feature of the navigation interface. The red line on the map shows where participants have been.

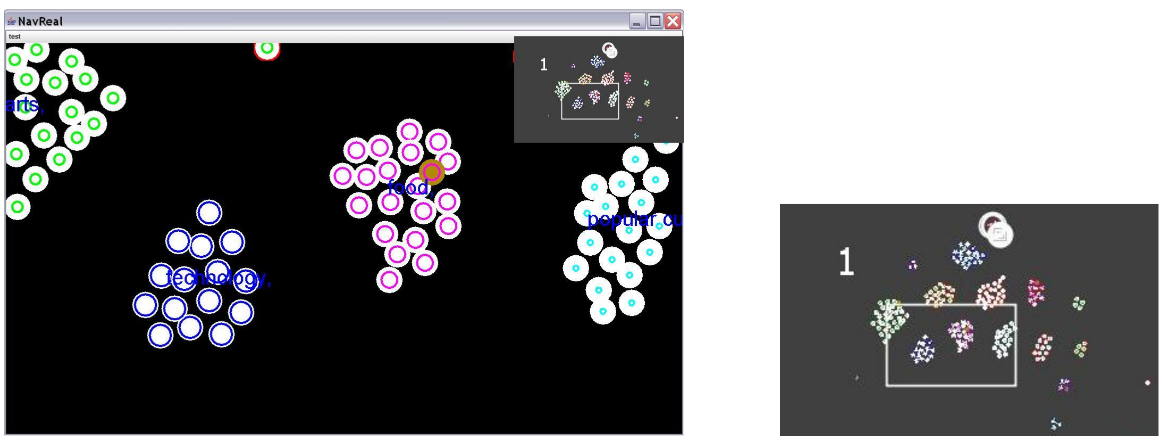


Figure 5.6: Navigation interface showing the prediction feature.

find out more, it is necessary to consider how much of a cluster has been listened to, as well as if it as been visited previously. When a prediction is asked for, a relevant cluster is predicted, and a random sound object is played from that cluster. If it is deemed a good prediction, users move to the predicted cluster by pressing "enter". If not, more predictions can be asked for. If a prediction is listened to for a reasonable amount of time, it is added to the *timeline*. A predicted cluster is marked with a circle in the map. Figure 5.6 shows the prediction feature of the navigation interface.

Listening Features

click-to-play/stop. This is the basic playback option. Click on the sound object to listen to it. To stop the sound, click on an empty part of the workspace. All sounds start from the beginning.

Audiosphere. The *audiosphere* features in the *ear* version of PAWS II. The *audiosphere* allows users to listen to more than one spoken word document at once. It is expected that although users are unable to attend to all sounds simultaneously (see Sect. 3.2), the separation of sounds in space, and different types of voices will mean that they will be able to focus on the appropriate sound while browsing. This is expected to reduce the time spent browsing for relevant spoken word documents. This function is toggled on and off using the right mouse button. The *audiosphere* is represented by an ear icon with a blue transparent circle (see Fig. 5.7). When the *audiosphere* hovers over an spoken word document, the content of that document is played. If it hovers over more than one document, all of the documents within the *audiosphere* will play. In order to make the sounds more intelligible to users, sounds are separated by distance and position in an audio space. This means that spoken word documents in the centre of the *audiosphere* are louder than those at the edge. To listen to a single spoken word document while in this mode, users click on the wanted document. This plays the document from the start, while stopping all other spoken word documents currently playing within the *audiosphere*.

Using a similar technique to the *audioView* in PAWS (see Sect. 4.2), 2D universe coordinates are transferred into 3D coordinates and mapped into a partial sphere using JOAL. In the *audioSphere* sounds are mapped to a round sphere, and the loudness of the sounds was adjusted using experimentation. To make enough of a difference in terms of loudness, the Z offset of the listener was changed to appear further away. This had the

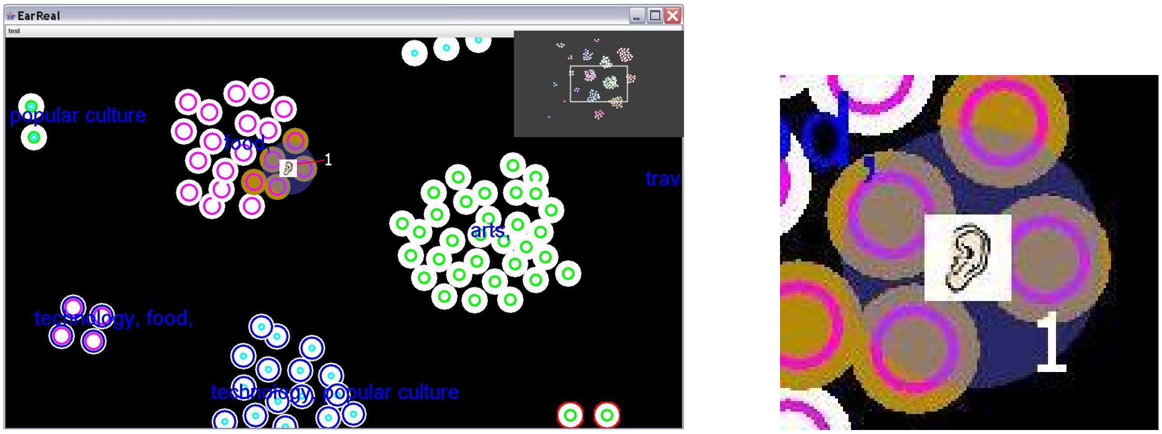


Figure 5.7: The *audiosphere* feature (represented by an ear icon and blue transparent bubble) show which spoken word documents are being listened to.

effect that the loudness and position changed more dramatically than if the listener was closer. Using the zoom feature did not affect the *audiosphere*.

5.3 Methodology

The experiment was a repeated-measures design for independent variables interface type with three levels (*basic, ear, navigation*) and seeking-task type with four levels (*find three, find as many as possible in three minutes, find specific object, comparison of two topics*). Participants carried out the experiment using all interfaces and seeking task types. Due to the risk of a learning effect, PAWS II interfaces were presented to participants in random order. The seeking-task type appeared in the same sequence each time (i.e., *find three objects, find as many objects in three minutes, find a specific object, comparison of two objects*). The audio excerpts and tasks used for the experiment assumed no prior knowledge. Participants were allocated to an interface type randomly.

5.3.1 Participants

In total, 18 participants took part. This was a different set of participants than took part in the organisation and re-find experiment. This was due to availability, and it gave the opportunity to study a different set of users who were not professionals within the area of knowledge management and, therefore, would have different experiences and comments. Of the 18, four worked professionally with music, and a further four took part in musical activities. Six non-native speakers (three Danish, one German, and one French) took part, the remainder were native English speakers. Four participants had experience with podcasting and two had experience with mindmapping tools. All participants took part in seeking tasks in their everyday life, and their level of English was sufficient for the tasks.

5.3.2 Apparatus and Materials

Questionnaires

Five questionnaires were used throughout the experiment. At the start, all participants completed a questionnaire to determine age group, native language, musical training, experience with spatial tools, and experience with collecting spoken audio (see App. C.4).

After each interface, another questionnaire was given, consisting of a set of five-point Likert questions [145] (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree) and open-ended questions. These were to gauge subjective satisfaction for each version (the questions are outlined in Tab. 5.3); (see App. C.5, C.6 and C.7).

Participants were asked to comment on the features within each version and describe their good points and possible improvements. Finally, they described what strategies they used to find relevant information, and provided comments that they felt related to the experiment, but had not been asked.

Question num.	**Question**
L1	It was easy to navigate around the information
L2	It was easy to find relevant objects
L3	The interface was easy to use
L4	There is too much information on the screen
L5	I enjoyed using this interface

Table 5.3: Table of Likert questions used in the seeking experiment.

Once participants had completed the experiment in all three interfaces of PAWS II, they were asked which had been the easiest and the most annoying interface out of the three, and for final comments on any aspect of the interface (see appendix C.8).

Audio Data

A database of 200 spoken word excerpts was collected from various podcast websites. The number was large enough to demonstrate the difficulties of organising and navigating within a spoken word collection. Excerpts ranged in length from 10 to 40 seconds. Each excerpt could belong to between 1 and 3 categories out of a possible 5 categories (travel, technology, arts, popular culture and food). A smaller database of 15 spoken word documents was used for the training period.

5.3.3 Procedure

Prior to the experiment, participants received training on the relevant interface of PAWS II. Instructions on paper explaining the different features were available, as well as a verbal walk-through of the relevant interfaces features (see App. C.1, C.2 and C.3). When participants were comfortable with the version, they completed a set of training tasks. This was to allow them to practice the tasks before they would be timed for the real experiment. It was also an opportunity for them to ask any questions regarding any aspect of the experiment, and to check the loudness of the audio output.

When the training was complete, the participants completed four tasks using the relevant interface. The four tasks were:

1. *Find three* podcasts related to X (a specified topic) This is an exhaustive search task that is restricted by the number that needs to be found.

2. *Find as many podcasts as possible* related to X (a specified topic) in three minutes. This is an exhaustive search task that is restricted by time.

3. *Find a specific* podcast. This is a known-item search task.

4. *Compare:* which has more podcasts, X or Y (two specified topics). This is an exploratory task.

The control panel for each task is shown in Figure 5.8. Number 1 shows a description of what information must be found and the tasks type. Number 2 shows when information has been selected. Number 3 shows the options available to participants. Clicking "Start" begins the experiment. Participants click this, once they understand what they need to find, and before they begin to listen to spoken word documents within the collection. Clicking "Add" adds the selected spoken word document to the list of answers. Clicking "Listen" plays the spoken word document that is selected within the control panel. Clicking "Remove" deletes a spoken word document from the answers. Once the required information has been found, or time has run out (as is the case in task 2), participants press "Complete".

Each task was a different type of finding task. In the first task, participants found three spoken word documents related to a topic. More may have existed, but they only needed to find three. This is an example of an exhaustive search task restricted by number to find. The search time, the number of spoken word documents found, the clusters looked in, and the path taken were recorded for future analysis.

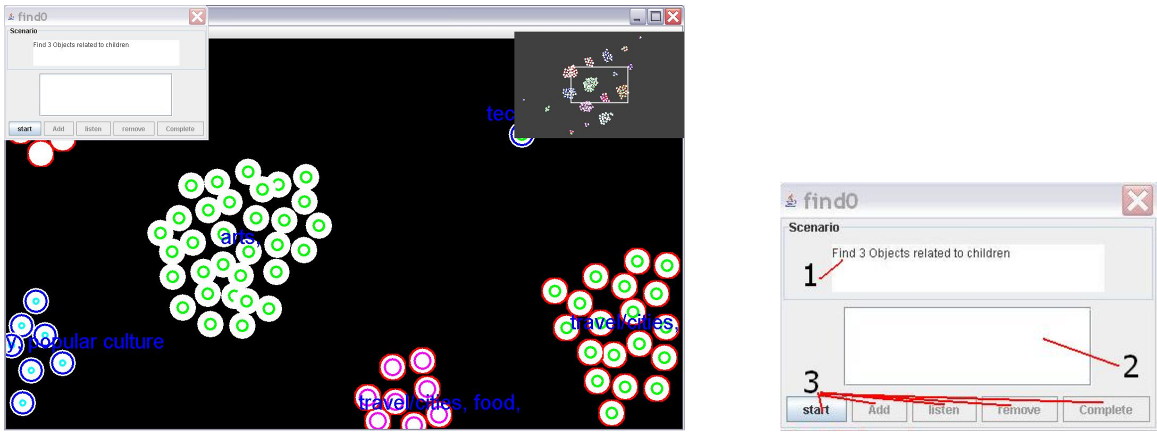


Figure 5.8: Control panel for seeking tasks.

The second task required finding as many as possible spoken word documents within three minutes. This is also an exhaustive search task, but restricted by time. Instead of focusing on finding highly relevant objects, anything that was related to the given topic could be selected. The focus here was to determine how much information could be found within the time limit. The number of spoken word documents found, clusters, and the path were recorded.

Task three asked for a specific spoken word document. This task had only one correct solution. The search time, spoken word document found, and the path were recorded. Finally, participants were asked to compare two topics to establish on which of the two more information existed. Participants could choose whether to browse first or whether to give an answer immediately. The search time and the number of spoken word documents listened to were recorded.

At the end of each interface, an interface specific questionnaire was administered (see App. C.5, C.6 and C.7). This was to determine the overall satisfaction with the interface. It also gave participants an opportunity to comment on aspects of the experiment. Finally, the participants described which interface had been the easiest and the most annoying to use, and they provided any additional comments that they felt would be useful for the analysis.

	Basic	Ear	Navigation
Time	3m 31s (SD 2m 26s)	3m 47s (SD 2m 9s)	3m 30s(SD 2m 16s)
Error rate	0.72 (SD 0.43)	0.76 (SD 0.27)	0.76 (SD 0.36)
Number listened to	39 (SD 32)	26(SD 27)	32(SD 26)
Average Time per object	5s (SD 2s)	4s (SD 3s)	4s(SD 2s)

Table 5.4: Results for find-three task. Time in seconds (s), error rate for the is percentage of correct items found. Average number of objects listened to and average time listened to for each object in seconds (s). Standard deviations are provided for all results.

5.4 Results

5.4.1 Task One - Find Three Relevant Objects

Participants were asked to find three relevant objects as quickly as possible. An answer was considered correct if a keyword relevant to the scenario was included in the sound object. Times, error rates, the number of objects listened to (including repeats), and average listening times are given in Table 5.4

A repeated-measure t-Test compared the *basic* with the *ear* interface and the *basic* with the *navigation* interface. No significant differences were found in the time, error rates, number of objects listened to, or length of time a sound object was listened to. The standard deviation is slightly higher for the *basic* version. This indicates more result variability.

5.4.2 Task Two - Find As Many Objects As Possible in Three Minutes

In this task, participants were asked to find as many related objects as possible within three minutes. An answer was considered correct if a keyword relating to the textual description was contained in the spoken word document. Table 5.5 shows the number of

	Basic	Ear	Navigation
Number found	3.28 (SD 2.52)	3.44 (SD 2.25)	4.22 (SD 2.84)
Number listened to	32 (SD 16)	22 (SD 17)	39(23)
Average time per object	7s (SD 4s)	4s (SD 3s)	7s (SD 4s)

Table 5.5: Results for find as many objects as possible in three minutes. Number found is the number of correct items found, given with standard deviations. Number of objects and average time listened to in seconds (s).

correct objects found with standard deviations, the number of objects listened to, and the average listening time for each object.

A repeated-measures t-Test was performed comparing the *basic* with the *ear* interface, and the *basic* with the *navigation* interface. No significant differences were found for the number of objects found in neither the *basic* - *ear* nor in the *basic* - *navigation* comparison. There was no significant difference in the time listened to for either comparison. The number of objects listened to was significantly greater in the *navigation* interface than in the *basic* interface ($t(17) = 0.52$, $p = 0.03$).

5.4.3 Task Three - Find Specific

Participants were asked to find one specific object. There was only one correct way of solving this task. Times, error rates, the number of objects listened to (including repeats), and average listening times per object were recorded (see Tab 5.6).

A repeated-measures t-Test was performed comparing the *basic* and the *ear* interface and the *basic* and the *navigation* interface. With the *ear* interface the specific objects were found significantly faster than with the *basic* interface ($t(17)= 2.48$, $p = 0.02$). The *navigation* interface produced faster results on average than the *basic* interface, but not significantly so. There was no significant difference in the number of objects listened to, or the average time each object was listened to for the different interfaces.

	Basic	**Ear**	**Navigation**
Time	3m 13s (SD 2m 38s)	1m 57s (SD 1m 42s)	2m 28s (SD 2m 3s)
Error rate	0.61 (SD 0.50)	0.67 (SD 0.49)	0.78 (SD 0.43)
Number listened to	36 (SD 34)	19 (SD 17)	29 (SD 25)
Average time per object	7s (SD 3s)	5s (SD 4s)	7s (SD 3s)

Table 5.6: Results from Task 3. Time given in seconds with standard deviation. Error rate for number of correct objects found, again presented with standard deviations. Number of objects listened to and average time per object in seconds.

	Basic	**Ear**	**Navigation**
Time	1m 33s (SD 1m 45s)	1m 33s (SD 1m 30s)	1m 3s (SD 1m 8s)
Error	0.83 (SD 0.38)	0.72 (SD 0.46)	0.78 (SD 0.43)
Number listened to	20 (SD 23)	10 (SD 12)	11 (SD 11)
Average time per object	18s (SD 18s)	11s (SD 12s)	15s (SD 12s)

Table 5.7: Results from Task Four. The time is given in seconds, the number of objects listened to is the average number that was listened to before giving a response, and the error is the percentage of correct answers for each interface. The number of objects listened to and the average listening time per object is also given. All results are given with a standard deviation.

5.4.4 Task Four - Comparison

Participants were asked to indicate which of two subjects had more objects related to it. It was possible to either give an answer immediately, or to browse within the collection first. Finding times and the number of objects listened to were recorded. Table 5.7 shows the time, number of objects listened to, the average listening time per object, and the error rate for each interface.

A repeated-measure t-Test was conducted comparing the *basic* and the *ear* interfaces, and the *basic* and the *navigation* interfaces. No significant differences were found. The

	Basic	**Ear**	**Navigation**
Zoom	3 (SD 3)	4 (SD 4)	3 (SD 3)
Workspace pan	29 (SD 24)	16 (SD 12)	17 (SD 10)
Map pan	1 (SD 3)	1 (SD 1)	1 (SD 3)
Map click	0 (SD 0)	0.3 (SD 1)	0.3 (SD 1)

Table 5.8: Navigation features used for the *basic, ear,* and *navigation* versions of PAWS II

high standard deviation for time was perhaps due to the strategy used to solve such a task. Two types of participants were apparent. Firstly, those who preferred to listen to the collection more before responding. Secondly, those who gave an answer without further listening to the collection. The number of objects listened to was lower in both the *ear* and the *navigation* version, indicating that more of the collection was explored and listened to prior to the final task.

5.4.5 General Navigation Patterns

The path participants took while completing the tasks was recorded for analysis. Data captured included the use of different features (visual, audio, and navigation) and sound object data (number of sound objects listened to, how long they were listened to, etc.).

The analysis of features showed that panning using the workspace was employed frequently. From observation, most participants used the zoom feature at the beginning of the experiment to get an overview of the information on the workspace. For the remainder of the experiment, the workspace was left at a zoomed out position. The map's pan and the click features were rarely used, and only by those participants who did not use the zoom feature. The features were employed in a similar way for all versions, except for the pan, which was used significantly more frequently in the *basic* version than in both the *ear* and the *navigation* versions. The results are summarised in Table 5.8

	Basic	**Ear**	**Navigation**
Total objects listened to	126 (SD 63)	76 (SD 50)	111 (SD 32)
Total unique objects listened to	85 (SD 34)	52 (SD 35)	78 (SD 15)
Total playing time	12m 33s (SD 5m 36s)	7m 25s (SD 6m 19s)	11m 11s (SD 3m 17s)
Average listening time (s)	6s (SD 2s)	5s (SD 3s)	6s (SD 2s)

Table 5.9: Number of sound objects listened to, number of unique objects listened to, total playing time (in minutes and seconds), and average listening time per object (in seconds). All presented with their standard deviations.

5.4.6 Listening Tactics

In order to know if the right information is found within audio, it is necessary to listen to the contents. The results presented here describe listening patterns derived from the recorded paths. The results for the *ear* version is for sound objects that have been deliberately listened to (i.e., clicked on). Hovering over multiple sound objects did not register as a result.

A repeated-measures t-Test was performed comparing the total playing time. The *basic* version was compared against both the *navigation* and *ear* versions. Participants spent significantly less time listening to sound objects in the *ear* version ($t(17) = 2.64$, $p = 0.02$). There was a significant difference between the total playing time spent when comparing the *basic* and *ear* versions. This indicates that the ability to listen to a number of sound objects simultaneously reduces the time needed to listen to a collection. The average listening time for all three versions was similar (basic: 6s; ear: 5s; and navigation: 6s).

	Basic	Ear	Navigation
Cluster, systematic	11	10	13
Cluster, semi-systematic	7	7	5
Random	0	1	0

Table 5.10: Strategies used in each interface when seeking information.

5.4.7 Finding Strategies

Participants were asked to describe their finding strategies. Three strategies emerged: cluster systematic, cluster semi-systematic and random. A cluster systematic strategy was when participants looked for a relevant cluster, and then searched through the cluster without missing any objects. A cluster semi-systematic is similar, but not all objects in a cluster were listened to. Finally, random finding occurred when participants were not sure where to look, and so listened to objects without applying a known pattern. Table 5.10 outlines the strategies used in each interface.

The most popular approach was finding the correct cluster, and then systematically searching through it for relevant sound objects. No correlations between the navigation style and task time were found.

5.4.8 Additional Comments

Feedback was gathered from participants regarding different aspects of the three interfaces, in the form of Likert questions, and general comments. The numbers are based on a 5-point Likert scale (from 1 = strongly disagree to 5 = strongly agree). The average result shows how well a question was answered.

Question	Bas	Ear	Nav
It was easy to navigate	3.94 (SD 1.05)	4.06 (SD 0.87)	4.22 (SD 0.94)
It was easy to find objects	3.11 (SD 9.63)	3.22 (SD 1.06)	3.06 (SD 0.94)
The interface was easy to use	4.44 (SD 0.71)	4.33 (SD 0.6)	4.11 (SD 1)
There is too much information	2.22 (SD 1.11)	2.17 (SD 1.2)	2.28 (SD 1.01)
I enjoyed using it	3.61 (SD 1.24)	4 (SD 1.24)	3.6 (SD 1)

Table 5.11: Results from Likert questions.

Likert Questions

General Likert questions were asked for each interface. The results are summarised in Table 5.11. No significant differences in satisfaction were found between the *basic* and the *ear*, and the *basic* and the *navigation* interfaces.

General Comments

The general comments from participants can be divided into eight categories: usability, visual elements, audio elements, organisation, *audioSphere*, *timeline*, and *prediction*.

Usability. Most participants considered all three interfaces easy to use. This is reflected in the Likert scores.

Visual elements. The visual elements that were appreciated were the zooming and panning of the workspace. The map offered a good overview, but few users used the click and the pan features available. *IListen* was appreciated, although a suggestion to improve this feature was to gradually change the background colour depending on how much of an sound object was listened to (i.e., a sound object that was listened to for a short period would be closer to white than one that had been listened to for a longer period). This would have been especially helpful in the *ear* interface where the *iListen* status was changed on hover. Many of the older participants (over 40's category) requested larger font sizes for labels. This did not seem to be an issue for the younger participants. The ability to manually re-label clusters was requested by some participants.

Audio elements. Participants liked the easy access to the audio content. Some preferred the *click-to-hear*, while others preferred *audioSphere* approach. Those preferring the *click-to-hear* felt that the number of objects played simultaneously in the *audioSphere* made it difficult to concentrate on any given object, and found it therefore more confusing. There were also additional difficulties when using *audioSphere* in a large cluster as the number of sounds playing was felt to be overwhelming. Comments for the *audiosphere* were mixed. Some preferred to search with the feature on, and only when a keyword had been found, listen to spoken word documents individually, while others found the simultaneous presentation of several spoken word documents too confusing to work with. In general, most participants felt that the maximum number of spoken word documents they could listen to at once was around three. This number depended on several factors, such as the position of the spoken word documents, the type of voice (male, female), tone of voice (assertive, passive), and accent. Many felt that with a longer training period, some of the experienced difficulties may be reduced.

Organisation. Most users provided comments on the organisation. Although many felt that a spatial organisation was easy to use, some wanted to be able to manually re-organise the collection for their own needs. Another suggestion was a dynamic re-organisation of the collection depending on the search criteria.

Timeline. The timeline was not used by participants. Reasons for this included the lack of a suitable task for it, that verifying information was not necessary, and that using a spatial visual feature, such as zooming or dragging was easier, than using the timeline. This concept is popular within browser-based organisations. Possibly because, in many cases, only a small part of the collection can be viewed at once. With PAWS, participants had an overview of the organisation of the collection at all times. When participants were asked why the feature was not used, they reported that the tasks given did not warrant the use of such a feature.

Prediction. The prediction was used as a last resort. If a spoken word document could not be found, then a participant may have tried to find it using the prediction feature. A few participants felt that it was a good idea, but most did not think it provided an accurate prediction, as they wanted to predict by object property (the ones they had been listening to) instead of the category their previous choices belonged to. Possible improvements for this feature include allowing participants to mark what sound objects they would want the prediction to be based on, and using content instead of category prediction. The *prediction* was used by most of the participants at least once. If a prediction proved successful, participants would try to use it more often. However, if it was unsuccessful, it was not used again. A few of the participants did not attempt to use the feature, stating that an algorithm was not able to help them in this way. Confidence in a feature to perform well appears to be a determining factor when choosing to use a feature or not.

5.5 Discussion

The main findings of this experiment were that the *audiosphere* significantly improved performance on known-item search task. The time spent listening to individual sound objects also significantly decreased with the *audioSphere*. This indicates that presenting several spoken word documents simultaneously improves performance when users know exactly what they are looking for. The feature may be more useful in a personal collection. In the organisation and re-find experiment, the time taken to re-find a known-item in a collection of 20 was on average 36 seconds for the extended audio version, and the collection needed to be organised first (on average 10 minutes). In PAWS II a similar known-item task took 1 minute 42 seconds in a collection of 200 objects. It is expected that the increase in size of a collection by a factor of 10, and the fact that the collection was unknown to the user, would play a significant role. However, it appears that under similar circumstances (i.e., both collections were the same size), the PAWS II *ear* version would fare better than PAWS I.

Again, it was highlighted that seeking tasks within a spoken word collection are performed significantly slower than in text-based collections. On average, participants listened to a sound object for five seconds before deciding if it was relevant. Scanning a text document for relevant details takes less time.

5.6 Chapter Summary

This chapter described PAWS II. In order to reduce the time taken to explore collections, different features were designed: *audiosphere, timeline, predictions,* and *ilisten*. These were combined with more traditional visualisation techniques of panning, zooming, and multiple views of information. The experiment confirmed that using *audiosphere* significantly reduced the time taken to complete known-item tasks, and reduced times for browsing tasks. Although the timeline and prediction features were used less, many participants agreed that the ideas seemed relevant, but that more development was required before they were usable. An analysis of the paths, showed that different strategies for seeking information were apparent. Common techniques were to first listen to spoken word documents randomly to get a mental model of the space prior to seeking, systematic seeking and semi-systematic seeking. In all cases, location-based finding was employed. This could have been due to the lack of availability of a query-based search engine.

Chapter 6

Conclusions

This work focused on supporting seeking tasks within spoken word collections. Advantages in gathering spoken word as an information resource have been known for some time, but the difficulty in accessing the spoken word within collections after it has been recorded meant that much of the gathered information has been transferred to text. Instead, the aim of this work was to maintain spoken word information in its audio form, and improve access to spoken word collections. A strength of this work was to focus on understanding and supporting the strategies and behaviours of individuals when performing seeking tasks, instead of replicating those strategies and behaviours. This lead to investigating ways in which to support organisation, navigation and seeking within collections. The approach that followed tackled the issues of supporting seeking tasks within spoken word collections by presenting the spoken word collection as a spatial organisation and visualisation, and including supportive playback and navigation features. Combining audio and visual modalities within the interface aimed to improve the accessibility of spoken word collections.

6.1 Contributions

The main findings of this research are as follows:

1. Simultaneous presentation of spoken word documents significantly improves performance in known-item search tasks.

2. Abstract audio cues that are used to describe properties within a spoken word document significantly decreased performance in known-item search tasks.

3. Organisation is an important sub-task in personal information management tools. The task of organising reduces the time spent looking for information. However, time spent to manually organise a collection is high. Therefore, many users may choose to not organise or maintain organisations.

4. Automatic organisations can be useful, providing adequate playback and navigation cues are available.

6.2 Research Questions

This section describes to what extent the research questions posed in Chapter 1.2 have been answered.

6.2.1 What Seeking Tasks and Strategies Need To Be Supported?

This research area was focused on identifying seeking tasks, and frequent strategies.

RQ1: Identify possible seeking tasks relevant to spoken word collections.

It was speculated that seeking tasks and strategies commonly employed for spoken word collections would be similar to those used in text and music collections. This was based on the assumptions that spoken word and text shared a lexical information, and that music and spoken word shared the medium it is presented in (i.e., audio). It was also assumed that seeking behaviours in collections would be similar regardless of the type of information being explored. The seeking tasks were generally described as either searching or browsing. More precisely, the tasks were identified as known-item, existence, exhaustive, and exploratory tasks. All tasks are applicable to both public and personal collections.

The study from Chapter 2 described common seeking tasks within personal music collections as either known-item or exploratory. A frequent task was to create a playlist; an exact match was not necessary. This indicated the usefulness of designing a tool that would support exploratory (browsing) tasks as well as known-item (searching) tasks. Exhaustive and existence seeking tasks were not described by participants in the study. This may be in part due to users being familiar with their collections.

RQ2: Identify possible seeking strategies frequently undertaken when seeking spoken word information.

Common strategies for known-item and exploratory seeking tasks include query-based searching, foraging, and location-based seeking. Query matching was not implemented in either prototype.

There are differences when seeking information in visual or audio collections. In spoken word collections, details such as tone of voice are difficult to search for. In both experiments, participants reported remembering different voices, especially if they were distinctive. In the second experiment, when asked to find as many spoken word documents related to Scotland as possible, native English speakers reported they listened to the speakers' accents. If a speaker appeared to be Scottish, they listened further to know if the spoken word document was relevant. Searching for audio properties is not possible with present query-based search engines. To support this type of searching, *multi-hear* and, later, *audiosphere* were designed to present several audio documents simultaneously. Both experiments demonstrated that spoken word collections much longer to search as long as they remain in audio form. A comparison of known-item search task in the organisation and re-find experiment was compared to a similar study for webpages, the time taken to find a correct spoken word document is approximately five times longer than that of a webpage.

Naturally, listening behaviours are an important aspect of seeking within audio collections. Observations in both experiments showed that participants listened briefly (approximately 5 seconds) to spoken word documents before deciding if they were relevant. This implies that if the wanted information is not heard in the first few seconds, the spoken word document may be discarded without further listening. Therefore, it may be assumed that the more patient users are, the more likely they are to find relevant information. This again highlights the need to explore methods to improve access to an individual spoken word documents.

In both experiments, three strategies of browsing were apparent.

1. Random selection of clusters before completing search tasks in order to gain content overview of the collection.

2. Jump to expected location, search systematically within cluster (similar to location-based searching).

3. Jump to expected location, listen to a few spoken word documents to decide on relevance. If nothing relevant is found, move to the next expected location.

The first strategy is similar to orienteering (described in Chapter 2). The second strategy closely matches teleporting (Chap: 2), and the third strategy is a combination of the two approaches. The choice of how to best approach a seeking task appeared to depend on the participant.

6.2.2 How To Support These Common Tasks Seeking Tasks and Strategies

Chapter 3 provided an overview of spatial,visualisation, and standard search tools and discussed different ways of employing them to support users' individual organisation strategies as well as browsing and searching type seeking tasks.

RQ3: How should a spoken word collection be organised?

Different types of users exist based on whether they organise their personal collections, and if they maintain these collections once organised. These observations were also found in studies on e-mail, electronic document, and physical document collections. For this work, these groups were known as *manual organisers*, *automatic organisers*, and *non-organisers*. In general, whether a collection was organised and maintained was a personal choice based on the perceived usefulness of the activity. Similar findings showed that when users organised a physical collection, spatial and emergent organisations were more common, but within electronic collections, classifications were used more. This is perhaps due to the file and folder system which is prevalent on most operating systems. The idea of this work was to provide users with an access to tools that support spatial organisation as this was expected to enhance their seeking abilities since spatial organisation is a skill they had developed over many years. Of the 36 participants, only one reported to preferring classifications and lists. All participants were able to use the tools successfully following a short training period. In the seeking experiment, some participants reported that the spatial layout gave additional information about the collection and encouraged them to browse more. The size of the clusters also indicated how long it would take to listen to a cluster.

The two experiments in this work considered two types of user groups: those who used manual and those who used automatic organisation. The organisation and re-find experiment began by participants organising a small collection of spoken word documents (20) and then completing re-find tasks. Participants completed this experiment twice (on different versions of PAWS). Participants took significantly longer to organise the collection in the second attempt than in the first. Re-find times were shorter in the second attempt than in the first, but not significantly so. This indicates that the time taken to organise a collection would affect the re-find time. A number of organisation strategies were employed to organise a collection of 20 spoken word documents. Three factors were found: use of space, alignment and organisation property. Again, this implies that organisation schemes are personal, and have developed over a period of time to suit a preferred seeking behaviour. Around half of the group decided to change their organisation strategy in the second attempt. This suggests that they felt that organisation was important, and that they had not reached an optimal strategy in the first attempt. However, the cost in terms of time to organise was high (around 9 minutes for 20 objects). For those users where organising is an effective stage in analysing data (such as in the case of sense-making tasks), this is perhaps a viable option. However, for those users with large collections, investing the time to organise is not realistic, and probably not desirable.

In the seeking experiment, the spoken word collection (200 objects) was organised automatically based on category similarity. However, at the end of the experiment, many of the participants commented that they would have liked the opportunity to re-organise part of the collection manually to suit their own needs. Another suggestion was to have dynamically changing views of the collection in order to highlight the relationships between the spoken word documents more effectively. In view of this, a combination of automatically and manually organisation techniques should be explored.

RQ4: What features support the different seeking tasks and strategies?

This research focused on supporting known-item and exhaustive search tasks. Instead of using query-based searching, the work focused on improving the accessibility of the collection, and in the case of PAWS strengthening memory cues. In the case of PAWS II, additional navigation and playback features were introduced to support location-based search strategies. As location-based searching had been highlighted as one of the most used strategies, the tools focused on improving this. Visualisations have been shown to support browsing effectively by providing different views of information. Therefore, this organisation style was chosen for PAWS and PAWS II.

In the case of re-find tasks, the most common strategy for locating information was location-based searching. This meant looking for information where it was expected to be. This strategy was common in both experiments. Query-based searching relied on users knowing the "right" information, and frequently this was not the case. In terms of spoken word documents, types of information that may be remembered are, amongst others, the tone of voice, and accent. These are difficult to search for with query-based searching. Finally, linear search as employed either once a suitable starting location had been found, or as a last resort.

For finding tasks, depending on how well the wanted information could be articulated, or if an exact file name was known, a query search provided an answer quickly. However, if an information need was more difficult to articulate, common strategies included looking in a collection that was thought to contain the answers, and refining the search from that point.

RQ5:How can the audio medium be used to best effect within a tool?

Being able to listen to the audio content quickly was highlighted in several studies as an important factor for retrieval within audio collections. Different audio features were implemented in both PAWS and PAWS II. In PAWS the use of abstract audio cues detracted from ability to re-find information successfully. This may have been due to learning factors. However, it may indicate that abstract audio has a learning cost, and users prefer to use visual cues. The most successful audio feature of PAWS was the multi-hear. It allowed users to listen to more than one document at once. This feature was refinded in PAWS II and results showed that users were significantly faster while using *audiosphere* for known-item seeking tasks. However, this was less effective for browsing type tasks, as the amount of audio information overloaded the user. For browsing tasks *hover-to-hear* provided better results.

Perhaps the use of abstract audio cues to describe audio information is less useful than using abstract audio cues to describe other forms of information, such as multi-variate statistical information. A further test using a different information type, such as highly complex visual data could clarify this issue further.

6.3 Limitations

A limitation for this research was the lack of a suitable specialist user group. These lead to a number of issues that affected the research: firstly, a lack of research that discussed seeking behaviours within spoken word documents: secondly, a lack of participants for the experiments;, finally, the inability to use observational techniques to fully understand how seeking was implemented in a collection.

In order to minimize the impact of a lack of prior research within spoken word collection seeking, two similar information collections, music and text, were used in order to give a reasonable starting point. Text and spoken word are similar in the way that information is presented through natural language. Music and spoken word are similar because they are both delivered in the audio medium. Therefore, it is expected that combining research in the areas of text and music seeking would provide a reasonable starting point for understanding seeking within spoken word collections. The tools discussed in

Chapter 3 focused for the most part on text and music retrieval. However, it is expected that many of the techniques described may be applied to spoken word collections. The number of users who have a spoken word collection is assumed to be greater now already than at the start of this research, so conducting similar experiments now and comparing them to the results from this work may show the development of strategies.

To minimize the impact of the lack of a user group, all studies used participants who were experienced in seeking within either text or music collections or both. It is anticipated that as the availability of spoken word collections increases, a user group will emerge. This user group will probably have prior experience of searching for both text and music documents, and the strategies and behaviours used will be based on their experiences of seeking within other types of collections. Naturally, it was not possible to investigate seeking within users' personal collections. Therefore, in both empirical experiments, the spoken word collections were unknown to all participants.

6.4 Future Work

A number of issues have been highlighted in this work. Firstly, the number and variety of strategies and behaviours observed in this study implies that tools that provide one style of organisation are unlikely to support all users in their tasks. Experiments comparing manual, automatic, and no organisation tools may help to provide recommendations on how information spaces may be best personalised towards users. Perhaps the use of better navigation support and better accessibility to audio may mean that one type of automatic organisation could be used.

Secondly, increasing the speed at which an individual spoken word document can be listened to may reduce the time taken to decide if a found file was relevant or not. On average, users listened to a file for five seconds before deciding on its relevance. In some cases, spoken word documents can last for hours. The fields of content extraction may provide possible solutions for this area.

Thirdly, the results provided in the experiments considered small and personal collections, but how could this be adapted to suit large public collections? At present, most spoken word collections are presented as classifications (usually by genre). Would the use of spatial representations improve browsing tasks? A recognised problem of spatial organisation is that as a collection increases, and the number of relationships between the individual spoken documents increase, the resulting visualisations may be too complex for users to interact with. Studies looking at large collections may provide answers to this issue.

Appendix A

Appendix - Organise and Re-find Experiment

A.1 Instructions for PAWS - Basic Audio

A.1.1 Features

Sound objects

A sound object (see Fig. A.1) contains a spoken word document. It can be moved either in the viewport or in the map, by clicking and dragging to a new position. To listen, either on the viewport or the map, click to select (icon will turn orange) and press play on the audio control. To stop, click to select then press stop. If a sound object is already playing when play is pressed for another sound object, the first sound object will stop.

Figure A.1: Icon of a sound object.

Landmarks

A landmark (see Fig. A.2) is a stationary object. It is represented by a blue square on the map.

Figure A.2: Icon of a Landmark object.

Workspace

The workspace is your working space. It shows part of the possible area (universe). To navigate, "click and drag" within the workspace. If the cursor is positioned on top of a sound object, the sound object will be repositioned. If not, the workspace will move.

Map

The map shows all of the possible workspace. Blue square represent landmarks, and orange squares represent sound objects. A green rectangle represents the workspace's position. To navigate, click and drag within the green rectangle to move the workspace's position, or click and drag on top of sound objects to move their position.

A.2 Instructions for PAWS - Extended Audio

A.2.1 Features

Sound objects

A sound object (see Fig. A.3) contains a spoken word document. It can be moved either in the workspace or in the map, by clicking and dragging to a new position. To listen, either on the workspace or the map, click to select (icon will turn orange) and press play on the audio control. To stop, click to select then press stop. If a sound object is already playing when play is pressed for another sound object, the first sound object will stop. It is also possible to listen to a sound object by hovering over it. The icon turns red in both the map and workspace when a sound is playing.

Figure A.3: Icon of a sound object.

Landmarks

A landmark (see Fig. A.4) is a stationary object. It is represented by a blue square on the map. To add a sound to a landmark, click on the landmark button in the audio control, and attach a sound to the landmark. To listen to the landmark, either hover over it, or select and press play on the control panel. Hovering over a landmark causes the nearest five sound objects to play. All objects playing are turned red in the viewport and map.

Figure A.4: Icon of a Landmark object.

Workspace

The workspace is your working space. It shows part of the possible area (universe). To navigate, "click and drag" within the workspace. If the cursor is positioned on top of a sound object, the sound object will be repositioned. If not, the workspace will move.

Map

The map shows all of the possible workspace. Blue squares represent landmarks, and orange squares represent sound objects. A green rectangle represents the workspace's position. To navigate, click and drag within the green rectangle to move the workspace's position, or click and drag on top of sound objects to move their position.

Audio Control

The audio control allows different parts of the sound object to be played. Content refers to only the spoken word to be played. For number, When there is only once voice in the excerpt, one voice plays, if there is more than one voice, a mixture of voices play simultaneously. For length, a short sound means the excerpt is short, a long sound means the excerpt is long. Click on the boxes to select which aspect to listen to. It is possible to listen to more than one.

A.3 Tasks for Organise and Re-find Experiment

A.3.1 Organise

- Click “Add it”. A new sound object will appear on the workspace. Listen to the spoken word document as often as necessary. Position accordingly. Repeat until all 20 sound objects have been added and organised on your workspace. You may listen and reposition other sound objects as required.

- Attention should be paid to both the content and audio properties of the sound object.

- You should organise in such a way that you will be able to re-find different sound objects quickly.

A.3.2 Re-find a Specific Sound Object

After listening to the wanted sound (as often as needed) find the sound in your workspace. Click to select and then click “Found” once you have found it. This task is repeated five times.

A.3.3 Re-find a Group of Related Objects, Topic and Sound Properties

Find as many sound objects as possible related to a written description. When a sound is related, click to select, and press “Add”. When you have found everything that is related, press “Complete”. This task is repeated five times.

A.4 Screen shots for Organisation Task

Grouped into sets, ordered by participants.

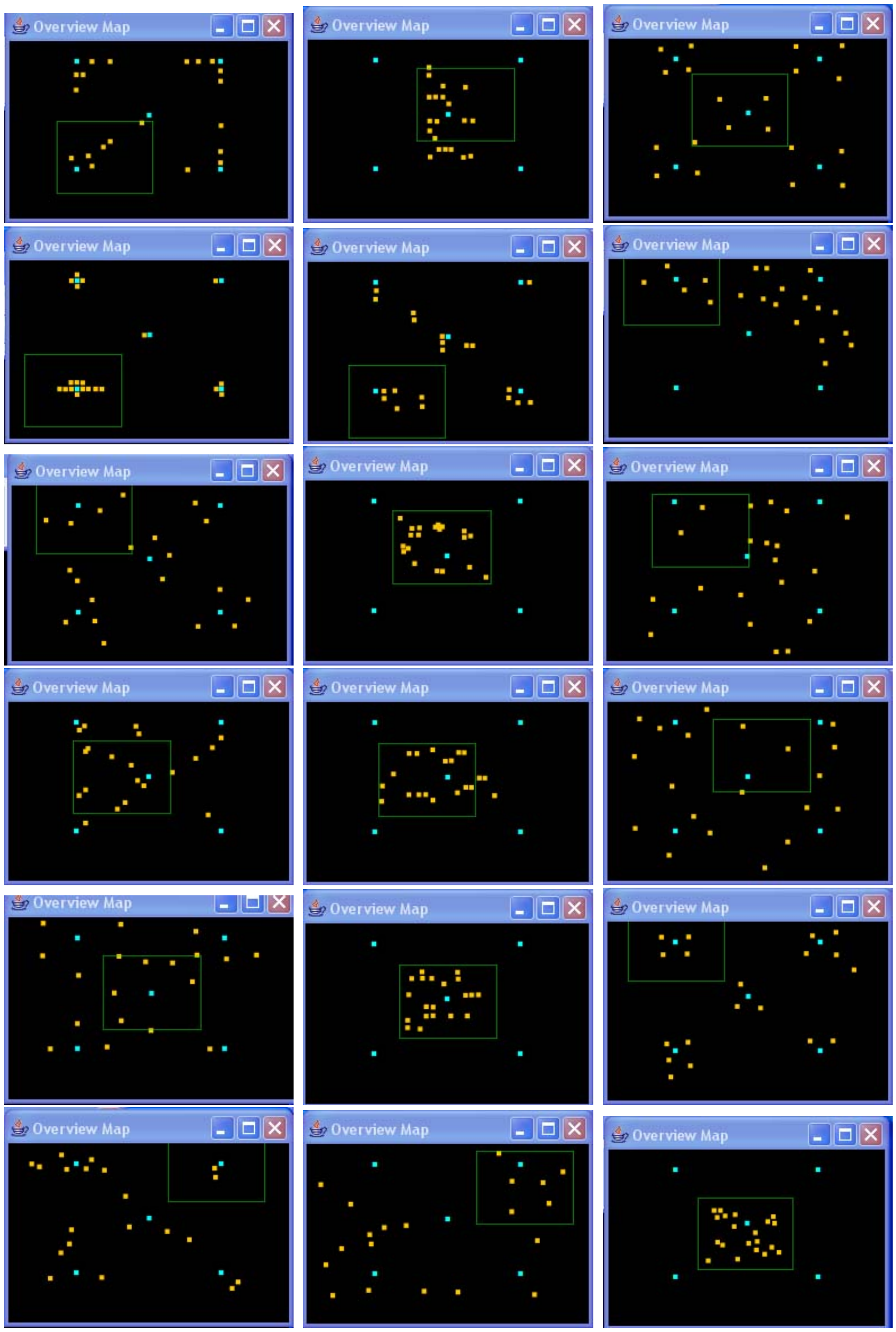


Figure A.5: Set 1 organisations

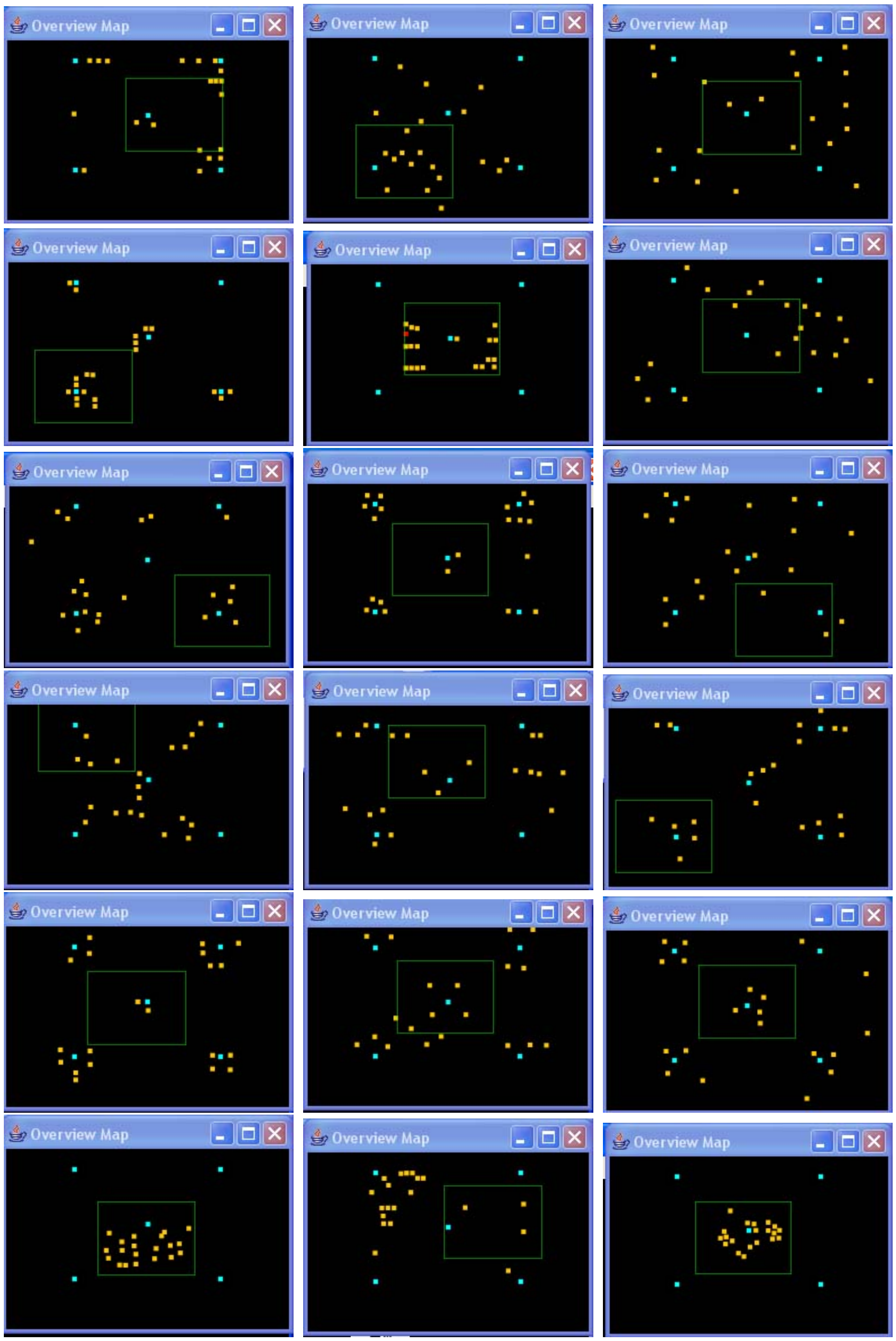


Figure A.6: Set 2 organisations.

Appendix B

Raw Data for Exp. 1

Participant	Order (basic/extended)	**Basic audio**	**Extended audio**
1	2/1	407	465
2	1/2	424	555
3	2/1	522	127
4	2/1	806	859
5	1/2	320	414
6	2/1	591	365
7	2/1	312	315
8	1/2	390	863
9	1/2	456	708
10	1/2	735	473
11	1/2	337	619
12	1/2	621	823
13	2/1	903	856
14	1/2	525	865
15	2/1	524	594
16	2/1	782	65
17	1/2	530	501
18	2/1	218	277

Table B.1: Times taken to organise 20 objects in both basic and audio versions. All results given in seconds.

Participant	**Basic audio**		**Extended audio**	
	Time	**Error**	**Time**	**Error**
1	14	1.00	17	1.00
2	20	1.00	32	1.00
3	33	1.00	70	1.00
4	18	1.00	44	1.00
5	39	1.00	22	1.00
6	31	1.00	43	1.00
7	32	1.00	97	1.00
8	17	1.00	38	1.00
9	44	1.00	54	1.00
10	24	1.00	36	1.00
11	19	1.00	25	1.00
12	25	1.00	26	1.00
13	26	1.00	38	1.00
14	23	1.00	22	1.00
15	15	1.00	21	1.00
16	38	1.00	33	1.00
17	51	1.00	35	1.00
18	16	1.00	22	1.00

Table B.2: Times and error rates from Task 1. Times in seconds and error rates in number correct.

Participant	**Basic audio**		**Extended audio**	
	Time (s)	**Num. found**	**Time(s)**	**Num. found**
1	105	5.00	41	4.00
2	24	3.33	89	2.67
3	142	4.00	226	2.33
4	149	2.33	121	2.67
5	48	3.67	119	2.67
6	86	3.33	112	3.67
7	59	4.00	65	3.33
8	45	5.70	115	3.00
9	48	4.67	67	3.33
10	55	4.33	51	3.67
11	61	4.33	124	4.67
12	22	4.00	94	3.00
13	308	4.00	92	3.00
14	17	4.33	83	3.67
15	73	3.67	49	4.00
16	173	5.33	105	2.67
17	87	3.33	81	2.00
18	44	3.33	94	1.33

Table B.3: Times and error rates from Task 2. Times in seconds and error rates in number correct.

Participant	**Basic audio**		**Extended audio**	
	Time (s)	**Num. found**	**Time (s)**	**Num. found**
1	133	5.00	99	5.00
2	32	3.00	66	4.00
3	163	6.50	102	4.50
4	45	9.00	22	4.50
5	132	3.00	72	3.50
6	21	4.50	81	4.50
7	98	2.50	99	3.00
8	99	3.50	90	3.00
9	74	5.00	148	1.50
10	94	4.00	96	3.00
11	86	3.00	119	8.00
12	41	3.50	93	3.50
13	397	6.50	257	7.00
14	24	2.50	96	3.50
15	120	2.50	108	4.00
16	204	5.00	126	4.00
17	67	2.00	83	4.50
18	90	2.50	45	5.50

Table B.4: Times and number found from Task 3. Times in seconds, and error rates in number correct.

Participant	Basic audio					Extended audio				
	L1	**L2**	**L3**	**L4**	**L5**	**L1**	**L2**	**L3**	**L4**	**L5**
1	3	4	3	1	4	2	4	4	1	4
2	5	3	2	2	4	5	3	2	2	4
3	2	3	2	2	5	4	2	4	2	4
4	5	4	4	2	4	4	3	4	2	4
5	5	3	4	1	4	3	4	4	2	4
6	3	3	3	1	2	1	4	5	1	4
7	2	1	3	1	4	4	3	5	1	5
8	4	3	4	1	5	3	4	4	1	4
9	2	4	2	1	4	2	3	3	3	3
10	4	5	3	2	4	5	5	5	2	5
11	4	3	3	2	5	4	4	3	2	4
12	3	3	4	2	3	3	3	3	2	4
13	3	4	3	1	2	4	4	3	2	4
14	2	3	5	1	3	4	3	4	1	3
15	4	4	4	1	4	5	4	4	1	5
16	2	2	2	1	2	2	2	3	3	4
17	2	3	4	1	4	2	2	4	4	5
18	4	4	3	1	3	4	4	4	1	4

Table B.5: Likert questions for re-find experiment.

Appendix C

Appendix - Seeking Experiment

C.1 Instructions UI B

- **Audio Positioning:** The volume of the sound depends on where the spoken word document you are listening to is in comparison to where you are in the workspace. A spoken word document playing outside the workspace will be quieter than one playing inside the workspace.

- **Colours in Spoken word documents:** The colours in the different spoken word document show what and how many categories the spoken word document belongs to. The background colour (*iListen*) of a spoken word document changes to show what has been listened to.

- **Clustering:** Spoken word documents are grouped by topic. If a spoken word document belongs in 2 categories, it will be grouped in a different cluster to that of spoken word documents that belong in one category or the other.

- **Zooming:** Use the mouse wheel to zoom in and out. It is necessary to click on the main view first before zooming will work.

- **To move workspace:** Either drag the white rectangle in the map, or click on blank space in the workspace and drag.

- **To listen to a spoken word document:** Move the mouse over a spoken word document and click.

- **To stop playback:** Click anywhere in the blank space of the workspace, or select another spoken word document to listen to.

C.2 Instructions UI E

- **Audio Positioning:** The volume of the sound depends on where the spoken word document you are listening to is in comparison to where you are in the workspace. A spoken word document playing outside the workspace will be quieter than one playing inside the workspace.

- **Colours in Spoken word documents:** The colours in the different spoken word document show what and how many categories the spoken word document belongs to. The background colour (*iListen*) of a spoken word document changes to show what has been listened to.

- **Clustering:** Spoken word documents are grouped by topic. If a spoken word document belongs in 2 categories, it will be grouped in a different cluster to that of spoken word documents that belong in one category or the other.

- **Zooming:** Use the mouse wheel to zoom in and out. It is necessary to click on the main view first before zooming will work.

- **To move workspace:** Either drag the white rectangle in the map, or click on blank space in the workspace and drag.

- **To listen to a spoken word document:** Move the mouse over a spoken word document and click.

- **To stop playback:** Click anywhere in the blank space of the workspace, or select another spoken word document to listen to.

- **"Audiosphere":** The *audiosphere* allows more than one spoken word document to be played at one time. This is represented visually by a blue bubble around the mouse pointer (which turns into an ear when activated). Spoken word documents closer to the centre of the ear are louder than those further away. Only spoken word documents within the blue radius are played. The *audiosphere* is only available in the main view. To activate and deactivate the *audiosphere* use the "right" button on the mouse.

C.3 Instructions UI N

- **Audio Positioning:** The volume of the sound depends on where the spoken word document you are listening to is in comparison to where you are in the workspace. A spoken word document playing outside the workspace will be quieter than one playing inside the workspace.

- **Colours in Spoken word documents:** The colours in the different spoken word document show what and how many categories the spoken word document belongs to. The background colour (*iListen*) of a spoken word document changes to show what has been listened to.

- **Clustering:** Spoken word documents are grouped by topic. If a spoken word document belongs in 2 categories, it will be grouped in a different cluster to that of spoken word documents that belong in one category or the other.

- **Zooming:** Use the mouse wheel to zoom in and out. It is necessary to click on the main view first before zooming will work.

- **To move workspace:** Either drag the white rectangle in the map, or click on blank space in the workspace and drag.

- **To listen to a spoken word document:** Move the mouse over a spoken word document and click.

- **To stop playback:** Click anywhere in the blank space of the workspace, or select another spoken word document to listen to.

- **Highlighting:** When you are exploring and before a decision is made, the map can show the different possibilities. When navigating, the current spoken word document is highlighted using a square marker and a red path shows the history. When exploring the predictions, a round marker is used. In general, when something is highlighted, pressing “enter” moves to that location on the screen.

- **Selection:** To select a spoken word document, click on it. This adds it to the history, and provides new predictions for it.

- **Back/Forward:** Press “$\leftarrow$” to move back. press “$\rightarrow$” to move forward. To move to that object press “Enter”

- **Predictions:** Predictions are based on previous choices. To get a predictions press “$\uparrow$”. The cluster the prediction comes from is marked with a circle on the mini-map. To move to the prediction press “enter”.

C.4 Demographic Questionnaire

1. **Name(For reference only):**
2. **age (circle appropriate):**
    - **Under 20**
    - **20 - 30**
    - **30 - 40**
    - **Over 40**
3. **Native language:**
4. **Level of musical training:**
5. **Level of experience with spatial organisation tools. If experience, what tools used?:**
6. **Level of experience with spoken audio (e.g. podcasts, radio interviews, etc). If experience, what tools used?:**

C.5 Questionnaire to UI B

C.5.1 Satisfaction

Answer the following questions using a scale between 1 - 5 (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree)

Question	1	2	3	4	5
It was easy to navigate around the information					
It was easy to find relevant objects					
The interface was easy to use					
There is too much information on the screen					
I enjoyed using this interface					

C.5.2 General

1. What were the good points of this interface?

2. What could be done to improve this interface?

3. Strategies used for finding relevant information.

4. Any further comments

C.6 Questionnaire to UI E

C.6.1 Satisfaction

Answer the following questions using a scale between 1 - 5 (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree)

Question	1	2	3	4	5
It was easy to navigate around the information					
It was easy to find relevant objects					
The interface was easy to use					
There is too much information on the screen					
I enjoyed using this interface					

C.6.2 General

1. When using the “audiosphere” there was too many sounds
2. The “audiosphere” was helpful
3. What were the good points of this interface?
4. What could be done to improve this interface?
5. Strategies for finding relevant information.
6. Any further comments

C.7 Questionnaire to UI N

C.7.1 Satisfaction

Answer the following questions using a scale between 1 - 5 (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree)

Question	1	2	3	4	5
It was easy to navigate around the information					
It was easy to find relevant objects					
The interface was easy to use					
There is too much information on the screen					
I enjoyed using this interface					

C.7.2 General

1. The back/forward navigation was helpful
2. The predictions were helpful
3. What were the good points of this interface?
4. What could be done to improve this interface?
5. Strategies used for finding relevant information.
6. Any further comments

C.8 Ending Questions

- Which interface was easiest to use? Why?

- Which interface was most annoying to use? Why

- Any additional comments:

Appendix D

Raw Data for Seeking Experiment

Participant	Basic		Ear		Navigation	
	Time (s)	Error	Time (s)	Error	Time (s)	Error
1	55	0	223	2	454	0
2	239	0	127	2	96	1
3	73	2	70	3	56	3
4	469	0	90	2	73	3
5	208	3	284	2	209	3
6	400	0	117	3	215	3
7	356	1	115	3	455	2
8	195	3	475	2	152	3
9	184	3	371	0	132	3
10	87	3	249	2	318	0
11	143	3	379	2	400	1
12	177	3	200	3	202	3
13	573	3	284	1	371	3
14	128	3	284	3	46	2
15	180	3	324	2	147	3
16	170	3	84	3	127	3
17	119	3	37	3	76	3
18	46	3	386	3	257	2

Table D.1: Task 1 time (in seconds) and error rate (number of correct objects found).

Participant	**Basic Error**	**Ear Error**	**Navigation Error**
1	2	0	3
2	2	1	4
3	3	7	1
4	9	2	7
5	5	5	1
6	8	1	2
7	2	1	11
8	0	5	7
9	2	3	3
10	6	7	9
11	2	3	3
12	3	6	4
13	2	5	3
14	2	2	3
15	2	1	7
16	0	3	1
17	6	6	4
18	3	4	3

Table D.2: Error rate (number of correct objects found).

Participant	**Basic**		**Ear**		**Navigation**	
	Time (s)	**Error**	**Time (s)**	**Error**	**Time (s)**	**Error**
1	290	0	235	0	182	0
2	12	0	79	0	229	0
3	98	0	46	1	414	0
4	181	1	83	1	9	1
5	504	1	386	0	322	0
6	298	1	107	1	220	1
7	152	1	0	0	110	1
8	202	0	103	0	335	1
9	225	0	126	1	226	1
10	149	0	60	0	14	1
11	260	1	243	1	78	1
12	46	1	12	1	13	1
13	567	1	176	1	62	1
14	65	1	40	1	45	1
15	42	0	236	1	133	1
16	309	1	110	1	173	1
17	35	1	6	1	26	1
18	53	1	65	1	69	1

Table D.3: Task 3 time (in seconds) and error rate (number of correct objects found).

Participant	Basic		Ear		Navigation	
	Time (s)	Error	Time (s)	Error	Time (s)	Error
1	257	1	210	0	41	1
2	123	1	156	1	192	1
3	210	1	17	1	5	1
4	12	1	56	0	99	0
5	64	1	129	1	14	0
6	267	1	196	0	164	0
7	8	1	9	0	7	1
8	30	0	155	1	8	1
9	6	1	318	1	196	1
10	6	1	150	1	54	1
11	119	1	15	1	115	1
12	319	0	13	1	96	1
13	16	0	19	1	8	1
14	107	1	67	1	100	1
15	124	1	60	0	8	1
16	9	1	83	1	10	0
17	3	1	4	1	5	1
18	6	1	8	1	9	1

Table D.4: Task 4 time (in seconds) and error rate (number of correct objects found, 1 = correct, 0 = wrong).

Participant	Basic					Ear					Navigation				
	L1	L2	L3	L4	L5	L1	L2	L3	L4	L5	L1	L2	L3	L4	L5
1	4	2	3	1	2	4	3	4	1	4	3	3	4	1	3
2	1	2	5	4	2	4	2	4	4	2	5	2	1	4	2
3	4	4	4	2	3	3	4	4	2	4	4	4	4	2	4
4	4	2	5	1	2	5	3	5	1	5	2	3	4	3	2
5	4	5	5	2	4	4	3	4	1	4	4	3	4	2	4
6	4	2	5	2	2	2	1	4	2	2	4	2	4	2	3
7	5	3	5	2	4	5	2	4	2	4	5	3	4	2	4
8	5	4	5	2	5	5	4	5	4	5	5	3	5	4	4
9	4	3	4	3	4	4	4	4	3	4	5	3	4	2	3
10	4	4	4	1	5	4	3	5	1	5	5	3	5	1	5
11	4	4	4	1	5	4	3	4	1	4	4	4	5	2	4
12	3	2	4	5	2	3	3	4	4	3	3	2	4	4	2
13	5	4	5	2	5	5	4	5	2	1	5	5	5	1	4
14	5	4	5	3	4	5	5	4	3	5	5	3	4	3	4
15	4	3	5	3	5	4	4	5	4	5	5	4	5	1	5
16	5	2	5	3	3	4	2	5	1	5	5	1	5	3	3
17	4	3	4	1	3	5	5	5	1	5	4	4	4	2	4
18	2	3	3	2	5	3	3	3	2	5	3	3	3	2	5

Table D.5: Likert questions for seeking experiment.

Bibliography

[1] BBC: BBC Four audio interviews. http://www.bbc.co.uk/bbcfour/audiointerviews/ (Accessed: 6th June 2007)

[2] BBC: BBC world service archive. http://news.bbc.co.uk/onthisday/ (Accessed: 6th June 2007)

[3] BBC: BBC world service archive. http://www.bbc.co.uk/worldservice/programmes/archive/ (Accessed: 6th June 2007)

[4] Thorsen, N.: Snart kan vi alle gå på skattejagt i DR's arkiver. http://politiken.dk/kultur/article314275.ece (29th May 2007)

[5] MySpace.com: Home page. http://www.myspace.com (Accessed: 14th May 2007)

[6] YouTube.com: Home page. http://youtube.com/ (Accessed: 14th May 2007)

[7] Flikr.com: Home page. http://www.flikr.com (Accessed: 14th May 2007)

[8] Gemmell, J., Bell, G., Lueder, R., Drucker, S., Wong, C.: MyLifeBits: fulfilling the Memex vision. In: Proceedings of the Tenth ACM International Conference on Multimedia. (2002) 235–238

[9] Trendwatching.com: Infolust. http://trendwatching.com/trends/infolust.htm **issue 35** (2006)

[10] Trendwatching.com: Generation C. http://trendwatching.com/trends/generation_c.htm **issue 23** (2005)

[11] Bush, V.: As we may think. Atlantic Monthly **176** (1945) 101–108

[12] Woods, D.D., Patterson, E., Roth, E.: Can we ever escape from data overload? a cognitive systems diagnosis. Cognition, Technology, and Work **4** (2002) 22–36

[13] Denning, P.J.: Infoglut. Communications of the ACM **49** (2006) 15–19

[14] Gustman, S., Soergel, D., Oard, D., Byrne, W., Picheny, M., Ramabhadran, B., Greenberg, D.: Supporting access to large digital oral history archives. In: Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries. (2002) 18–27

[15] Klemmer, S.R., Graham, J., Wolff, G.J., Landay, J.A.: Books with voices: paper transcripts as a physical interface to oral histories. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (2003) 89–96

[16] Susini, P., Vieillard, S., Deruty, E., Smith, B., Marin, C.: Sound navigation: sonified hyperlinks. In: Proceedings of the 2002 International Conference on Auditory Display. (2002)

[17] Tucker, R.C.F., Hickey, M., Haddock, N.: Speech-as-data technologies for personal information devices. Personal Ubiquitous Computing **7** (2003) 22–29

[18] Sawhney, N., Schmandt, C.: Nomadic Radio: Speech and audio interaction for contextual messaging in nomadic environments. ACM Transactions on Computer-Human Interaction **7** (2000) 353–383

[19] Foote, J.: An overview of audio information retrieval. Multimedia Systems **7** (1999) 2–10

[20] Tombros, T., Crestani, F.: User's perception of relevance of spoken documents. Journal of the American Society of Information Science **51** (2000) 929–939

[21] Munteanu, C., Baecker, R., Penn, G., Toms, E., James, D.: The effect of speech recognition accuracy rates on the usefulness and usability of webcast archives. In: Proceedings of the Conference of Computer Human Interaction. (2006)

[22] Gauvain, J.L., Lamel, L., Adda, G.: Transcribing broadcast news for audio and video indexing. Communications of the ACM **43** (2000) 64–70

[23] Nakatani, C.H., Whittaker, S., Hirschberg, J.: Now you hear it, now you don't: empirical studies of audio browsing behavior. In: Proceedings of the Fifth International Conference on Spoken Language Processing. (1998)

[24] Campbell, G.: There's something in the air - podcasting in education. Educase review (November/December 2005) 33–46

[25] Smith, S.: Radio free: enterprise: podcasting helps companies communicate. EContent (2005)

[26] Brown, E., Srinivasan, S., Coden, A., Ponceleon, D., Cooper, J., Amir, A., Pieper, J.: Towards speech as a knowledge resource. In: Proceedings of the Tenth International Conference on Information and Knowledge Management. (2001) 526–528

[27] Ellis, D.P., Lee, K.: Minimal-impact audio-based personal archives. In: Proceedings of the the First ACM Workshop on Continuous Archival and Retrieval of Personal Experiences. (2004) 39–47

[28] Vemuri, S., Schmandt, C., Bender, W., Tellex, S., Lassey, B.: An audio-based personal memory aid. In: Ubicomp. (2004) 400–417

[29] Skype.com: Home page. `http://www.skype.com` (Accessed: 14th May 2007)

[30] KishKish Sam: Home page. http://www.kishkish.com/sam (Accessed: 14th May 2007)

[31] HotRecorder: Home page. http://www.hotrecorder.com (Accessed: 14th May 2007)

[32] Audio Blogging: Home page. http://en.wikipedia.org/wiki/Audio_blog (Accessed: 14th May 2007)

[33] Winer, D.: Rss 2.0. http://cyber.law.harvard.edu/rss/rss.html (Accessed: 31 July 2007)

[34] Podcasting: Home page. http://en.wikipedia.org/wiki/Podcasting (Accessed: 14th May 2007)

[35] Hansen, J.H.L., Deller, J.R., Seadle, M.S.: Transcript-free search of audio archives for the National Gallery of the Spoken Word. In: Proceedings of the First ACM/IEEE-CS Joint Conference on Digital Libraries. (2001) 235–236

[36] Arons, B.: Speechskimmer: a system for interactively skimming recorded speech. ACM Transactions Computer-Human Interaction **4** (1997) 3–38

[37] Hirschberg, J., Whittaker, S., Hindle, D., Pereira, F., Singhal, A.: Finding information in audio: A new paradigm for audio browsing and retrieval. In: Proceedings of the ESCA ETRW Workshop. (1999)

[38] Furui, S.: Automatic speech recognition and its application to information extraction. In: Proceedings of the Thirty-seventh Conference on Association for Computational Linguistics. (1999) 11–20

[39] Shneiderman, B.: The limits of speech recognition. Communications of the ACM **43** (2000) 63–65

[40] TVEyes.com: Home page. http://www.tveyes.com (Accessed: 14th May 2007)

[41] Podzinger.com: Home page. http://www.podzinger.com (Accessed: 14th May 2007)

[42] Lee, J.H., Downie, J.S.: Survey of music information needs, uses, and seeking behaviours: preliminary findings. In: Proceedings of the Fifth International Conference on Music Information Retrieval. (2004) 441–446

[43] Dansk Folkemindesamling: Home page. www.dafos.dk (Accessed: 6th August 2007)

[44] Elsweiler, D., Ruthven, I.: Towards task-based personal information management evaluations. In: Proceedings of the Thirtieth Annual ACM Conference on Research and Development in Information Retrieval. (2007)

[45] Capra, R.G., Pérez-Quiñones, M.A.: Re-finding found things: Study of how users re-find information. Technical report, Virginia Tech (2003)

[46] Bruce, H.: Personal, anticipated information need. Information Research **10** (2005)

[47] Byström, K., Järvelin, K.: Task complexity affects information seeking and use. Information Processing and Management: an International Journal **31** (1995) 191–213

[48] Catledge, L.D., Pitkow, J.E.: Characterizing browsing strategies in the world-wide web. In: Proceedings of the Third International World-Wide Web Conference on Technology, Tools and Applications. (1995) 1065–1073

[49] Broder, A.: A taxonomy of web search. SIGIR Forum **36** (2002) 3–10

[50] Shneiderman, B., Byrd, D., Croft, W.B.: Clarifying search: A user-interface framework for text searches. Available at: http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html (1997)

[51] Järvelin, K., Ingwersen, P.: Information seeking research needs extension towards tasks and technology. Information Research **10** (2004)

[52] Kellar, M., Watters, C., Shepherd, M.: A field study characterizing Web-based information seeking tasks. Technical report, Faculty of Computer Science, Dalhousie University (2005)

[53] Wilson, T.D.: On user studies and information needs. Journal of Documentation **37** (1981)

[54] van Rijsbergen, C.J.: Information Retrieval. London: Butterworths. (1979)

[55] Bates, M.J.: The design of browsing and berry picking techniques for the online search interface. Online Review **13** (1989) 407–424

[56] Bergman, O., Beyth-Marom, R., Nachmias, R.: The user subjective approach to personal information management systems. Journal of the American Society for Information Science **54** (2003) 872 – 878

[57] Cutrell, E., Dumais, S.T.: Exploring personal information. Communications of the ACM **49** (2006) 50–51

[58] Berners-lee, T., Cailliau, R., Pollermann, B.: World-wide web: The information universe. Electronic Networking: Research, Applications and Policy (1992)

[59] Rosenfeld, L., Morville, P.: Information Architecture for the World Wide Web. second edn. O'Reilly Media, Inc. (2002)

[60] Bowker, G.C., Star, S.L.: Sorting things out: Classification and its consequences. MIT Press (1999)

[61] Bälter, O.: Keystroke level analysis of email message organization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (2000) 105–112

[62] Jacob, E.: The everyday world of work: Two approaches to the investigation of classification in context. Journal of Documentation **57** (2001)

[63] Nielsen, J.: The death of file systems. `http://www.useit.com/papers/filedeath.html` (February 1996: Accessed 21st June 2007)

[64] Dourish, P., Edwards, W.K., LaMarca, A., Salisbury, M.: Presto: an experimental architecture for fluid interactive document spaces. ACM Transactions on Computer-Human Interaction **6** (1999) 133–161

[65] Boardman, R.: Improving Tool Support for Personal Information Management. PhD thesis, Imperial College, London (2004)

[66] Card, S.K., Mackinlay, J.D., Shneiderman, B.: Readings in Information Visualization: Using Vision to Think. Academic Press (1999)

[67] Shipman, F.M., Marshall, C.C.: Spatial hypertext: an alternative to navigational and semantic links. ACM Computer Survey **31** (1999) 14

[68] Shipman, F.M., Hsieh, H., Maloor, P., Moore, J.M.: The Visual Knowledge Builder: a second generation spatial hypertext. In: Proceedings of the Twelfth ACM Conference on Hypertext and Hypermedia. (2001) 113–122

[69] Whittaker, S., Sidner, C.: Email overload: exploring personal information management of email. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (1996) 276–283

[70] Berlin, L.M., Jeffries, R., O'Day, V.L., Paepcke, A., Wharton, C.: Where did you put it? issues in the design and use of a group memory. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (1993) 23–30

[71] Malone, T.W.: How do people organize their desks?: Implications for the design of office information systems. ACM Transactions Information Systems **1** (1983) 99–112

[72] Boardman, R., Sasse, M.A.: "stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (2004) 583–590

[73] Bruce, H., Jones, W., Dumais, S.: Information behaviour that keeps found things found. Information Research **10** (2004)

[74] Kwasnik, B.: How a personal document's intended use or purpose affects its classification in an office. In: Proceedings of the Twelfth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. (1989) 207–210

[75] Barreau, D., Nardi, B.A.: Finding and reminding: file organization from the desktop. ACM SIGCHI Bulletin **27** (1995) 39–43

[76] Fertig, S., Freeman, E., Gelernter, D.: "Finding and reminding" reconsidered. SIGCHI Bulletin **28** (1996) 66–69

[77] Sutcliffe, A.: Multimedia and Virtual Reality: Designing Multisensory User Interfaces. Lawrence Erlbaum Associates (2003)

[78] Heller, R.S., Martin, C.D., Haneef, N., Gievska-Krliu, S.: Using a theoretical multimedia taxonomy framework. Journal on Educational Resources in Computing **1** (2001) 6

[79] Cole, I.: Human aspects of office filing: Implications for the electronic office. In: Proceedings of Human Factors Society. (1982)

[80] Nardi, B., Anderson, K., Erickson, T.: Filing and finding computer files. Technical Report Technical Report no. 118, Cupertino: Apple Computer, Inc. (1994)

[81] Hertzum, M.: Six roles of documents in professionals' work. In: Proceedings of the Sixth European Conference on Computer Supported Cooperative Work. (1999) 41–60

[82] Gwizdka, J.: Timely reminders: a case study of temporal guidance in pim and email tools usage. In: Extended Abstracts on Human Factors in Computing Systems. (2000) 163–164

[83] Cunningham, S.J., Jones, M., Jones, S.: Organising digital music for use: An examination of personal music collections. In: Proceedings for the International Symposium on Music Information Retrieval. (2004)

[84] Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings of the 1996 IEEE Symposium on Visual Languages. (1996) 336

[85] Teevan, J., Alvarado, C., Ackerman, M.S., Karger, D.R.: The perfect search engine is not enough: a study of orienteering behavior in directed search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (2004) 415–422

[86] Herder, E., Juvina, I.: Discovery of individual user navigation styles. In: Workshop on Individual Differences in Adaptive Hypermedia. (2004) 40–49

[87] Fidel, R., Efthimiadis, E.: Web searching behavior of aerospace engineers. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. (1999) 319–320

[88] Laplante, A., Downie, J.S.: Everyday life music information-seeking behaviour of young adults. In: Proceedings of the Seventh International Conference on Music Information Retrieval. (2006) 381–382

[89] Lansdale, M.: The psychology of personal information management. Applied Ergonomics **19** (1988) 55–66

[90] Abrams, D., Baecker, R., Chignell, M.: Information archiving with bookmarks: personal web space construction and organization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (1998) 41–48

[91] Elsweiler, D., Ruthven, I., Jones, C.: Towards memory supporting personal information management tools. Journal of the American Society for Information Science and Technology **58** (2007) 924–946

[92] Voicemail: Voicemail description. `http://en.wikipedia.org/wiki/Voicemail` (Accessed:2nd August 2007)

[93] Aguilar, F.J.: Scanning the Business Environment. New York: Macmillan (1967)

[94] Marchionini, G.M. In: Information Seeking in Electronic Environments. Cambridge University Press (1995) 106

[95] Wilson, T.D.: Information behaviour: An interdisciplinary perspective. Information Processing & Management **33** (1997) 551 – 572

[96] Choo, C.W., Detlor, B., Turnbull, D.: Information seeking on the Web: An integrated model of browsing and searching. First Monday: Available at: `http://firstmonday.org/issues/issue5_2/choo/index.html` **5** (2000)

[97] Blanc-Brude, T., Scapin, D.L.: What do people recall about their documents?: implications for desktop search tools. In: Proceedings of the Twelfth International Conference on Intelligent User Interfaces. (2007) 102–111

[98] Nuendo: Home page. http://www.steinberg.net/89_1.html (Accessed: 18th June 2007)

[99] Winamp Media player: Home page. http://www.winamp.com (accessed: 11th June 2007)

[100] Windows Media Player: Home page. http://www.microsoft.com/windows/windowsmedia/default.mspx (accessed: 11th June 2007)

[101] iTunes: Home page. http://www.apple.com/itunes/ (accessed: 11th June 2007)

[102] ID3 Tags: Home page. http://www.id3.org/ (accessed: 11th June 2007)

[103] Norman, D.A. In: Things that make use smart: defending human attributes in the age of the machine. Basic Books (1993) 175 – 180

[104] Clark, H., Clark, E.: Psychology and language: An introduction to psycholinguistics. Harcourt Brace Jovanovich (1977)

[105] Sachs, J.: Recognition memory for syntactic and semantic aspects of connected discourse. Perception Psychophysics **2** (1967) 437–42

[106] Rubin, D.: Very long-memory for prose and verse. Journal of Verbal Learning and Verbal Behavior **J 16** (1977) 611–621

[107] Rubin, D., Wenzel, A.: One hundred years of forgetting: a quantitative description of retention. Psychological Bulletin **103** (1996) 734–760

[108] Bower, G., Thompson-Schill, S., Tulving, E.: Reducing retroactive interference: An interference analysis. Journal of Experimental Psychology: Learning, Memory and Cognition **20** (1994) 51–66

[109] Halasz, F.G., Moran, T.P., Trigg, R.H.: Notecards in a nutshell. In: Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface. (1987) 45–52

[110] Marshall, C.C., Shipman, F.M., Coombs, J.H.: VIKI: spatial hypertext supporting emergent structure. In: Proceedings of the 1994 ACM European Conference on Hypermedia Technology. (1994) 13–23

[111] Fernández, L., Sánchez, J.A., García, A.: MiBiblio: personal spaces in a digital library universe. In: Proceedings of the Fifth ACM Conference on Digital Libraries. (2000) 232–233

[112] Buchanan, G., Blandford, A., Thimbleby, H., Jones, M.: Integrating information seeking and structuring: exploring the role of spatial hypertext in a digital library. In: Proceedings of the Fifteenth ACM Conference on Hypertext and Hypermedia. (2004) 225–234

[113] Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D., van Dantzich, M.: Data Mountain: using spatial memory for document management. In: Proceedings of the Eleventh Annual ACM Symposium on User Interface Software and Technology. (1998) 153–162

[114] Cockburn, A., McKenzie, B.: 3D or not 3D?: evaluating the effect of the third dimension in a document management system. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (2001) 434–441

[115] Cockburn, A., McKenzie, B.: Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (2002) 203–210

[116] Furnas, G.W.: Generalized fisheye views. In: Proceedings of CHI. (1986) 16–23

[117] Card, S.K., Robertson, G.G., York, W.: The webbook and the web forager: an information workspace for the world-wide web. In: CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems. (1996) 111–ff.

[118] McGookin, D.K., Brewster, S.A.: Dolphin: the design and initial evaluation of multimodal focus and context. In: Proceedings of the 2002 International Conference on Auditory Display. (2002)

[119] Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: A review. ACM Computing Survey **31** (1999) 264 – 323

[120] Pampalk, E.: Islands of music (2001)

[121] Neumayer, R., Lidy, T., Rauber, A.: Content-based organization of digital audio collections. In: Proceedings of the Fifth Open Workshop of MUSICNETWORK. (2005)

[122] Kohonen, T.: Self-Organizing Maps. 3rd edition edn. Springer (2001)

[123] Andrews, K.: Visualising information structures: Aspects of information visualisation. Professional Thesis. IICM, Graz University of Technology, Austria (2002)

[124] Brazil, E., Fernström, M., Tzanetakis, G., Cook, P.: Enhancing sonic browsing using audio information retrieval. In: Proceedings of the International Conference on Auditory Displays. (2002)

[125] Adamczyk, P.D.: Seeing sounds: exploring musical social networks. In: Proceedings of the Twelfth Annual ACM International Conference on Multimedia. (2004) 512–515

[126] Fernström, M., Brazil, E.: Sonic browsing: an auditory tool for multimedia asset management. In: Proceedings of the 2001 International Conference on Auditory Display. (2001)

[127] Jones, W., Dumais, S., Bruce, H.: Once found, what then? a study of "keeping" behaviors in the personal use of web information. In: Proceedings of the American Society for Information Science and Technology. (2002) 391–402

[128] Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., Robbins, D.C.: Stuff I've Seen: a system for personal information retrieval and re-use. In: Proceedings of the 26th annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval. (2003) 72–79

[129] Adar, E., Kargar, D., Stein, L.A.: Haystack: per-user information environments. In: Proceedings of the Eighth International Conference on Information and Knowledge Management. (1999) 413–422

[130] Clusty.com: Home page. `www.clusty.com` (accessed: 6th August 2007)

[131] Van Thong, J.M., Moreno, P.J., Logan, B., Fidler, B., Maffey, K., Moores, M.: SPEECHBOT: An experimental speech-based search engine for multimedia content on the web. Technical report, HP Labs (2001)

[132] Brungart, D.S., Simpson, B.D., Ericson, M.A., Scott, K.R.: Informational and energetic masking effects in the perception of multiple simultaneous talkers. The Journal of the Acoustical Society of America **110** (2001) 2527–2538

[133] Sawhney, N., Murphy, A.: ESPACE 2: an experimental hyperaudio environment. In: Conference Companion on Human Factors in Computing Systems. (1996) 105–106

[134] Schmandt, C.: Audio Hallway: a virtual acoustic environment for browsing. In: Proceedings of the Eleventh Annual ACM Symposium on User Interface Software and Technology. (1998) 163–170

[135] Morley, S., Petrie, H., O'Neill, A.M., McNally, P.: Auditory navigation in hyperspace: design and evaluation of a non-visual hypermedia system for blind users. In: Proceedings of the Third International ACM Conference on Assistive Technologies. (1998) 100–107

[136] Barbará, D.: The Audioweb. In: Proceedings of the Sixth International Conference on Information and Knowledge Management. (1997) 97–104

[137] Arons, B.: Hyperspeech: navigating in speech-only hypermedia. In: Proceedings of the Third Annual ACM Conference on Hypertext. (1991) 133–146

[138] Bly, S.: Presenting information in sound. In: Proceedings of the 1982 Conference on Human Factors in Computing Systems, ACM Press (1982) 371 – 375

[139] McGookin, D.K., Brewster, S.A.: Fishears: the design of a multimodal focus and context system. In: Proceedings of the IHM HCI. (2001)

[140] Blattner, M.M., Sumikawa, D.A., Greenberg, R.M.: Earcons and icons: Their structure and common design principles. Human-Computer Interaction **4** (1989) 11–44

[141] Gaver, W.W.: Auditory icons: Using sound in computer interfaces. Human-Computer Interaction **2** (1986) 167–177

[142] JOAL: Joal. `http://www.joal.org` (Accessed: June 2004)

[143] openAL: openal. `http://www.openal.org` (Accessed: June 2004)

[144] The VARK questionnaire: English questionnaire. `http://www.vark-learn.com/english/page.asp?p=questionnaire` (Accessed: June 2004)

[145] Likert, R.: A technique for the measurement of attitudes. Archives of Psychology **140** (1932)

[146] Gosset, W.S.: The probable error of a mean. Biometrika **6** (1908) 1–25

[147] Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics **1** (1945) 80–83

[148] Gröhn, M., Lokki, T., Takala, T.: Comparison of auditory, visual, and audio-visual navigation in 3D space. In: Proceedings of the 2003 International Conference on Auditory Display. (2003)

ISBN 978-87-7606-029-9